

UNIVERSITA' DI PISA
FACOLTA' DI ECONOMIA

CORSO DI LAUREA SPECIALISTICA
IN
STRATEGIA E GOVERNO DELL'AZIENDA

TESI DI LAUREA
IN
STATISTICA PER LE RICERCHE SPERIMENTALI E DI MERCATO

L'ANALISI STATISTICA DI DATI AZIENDALI

ESTRATTI DA BANCHE DATI:

LA PUBBLICITÀ PER I MasterDEA

Relatore:

Chiar.mo Prof. FRANCO MARTORANA

Correlatore:

Dott. FRANCESCO CIATARA

Laureando

ANTONIO DI PIPPO

Anno Accademico 2005/2006

Ringraziamenti

I contenuti di questo lavoro sono il frutto di un impegno congiunto di coloro che hanno contribuito direttamente o indirettamente alla sua realizzazione e alla sua stesura.

Anzitutto esprimo tutta la mia riconoscenza al Direttore del Dipartimento di Economia Aziendale, il Prof. Luciano Marchi, per avermi dato libero accesso agli uffici e ai dati sulla campagna pubblicitaria MasterDEA. Ringrazio, quindi, il Dott. Giuseppe D’Onza, per i suoi validi consigli, e la responsabile delle attività di marketing e delle relazioni esterne “MasterDEA”, la Dott.ssa Lisa Pampana, per la sua disponibilità e per avermi aiutato nella ricerca di informazioni rilevanti ai fini dell’analisi qui condotta.

Un ringraziamento particolare va al mio relatore il Prof. Franco Martorana, e al mio correlatore il Dott. Francesco Ciatara per gli utili suggerimenti e per la loro valida collaborazione nella redazione e revisione del presente lavoro.

Per la disponibilità e le attenzioni dedicate ai propri studenti ringrazio tutti i docenti del corso di laurea specialistica in “Strategia e governo dell’azienda” ed in particolar modo il Presidente del corso Prof. Silvio Bianchi Martini.

Da ultimo non potrò mai ringraziare abbastanza la mia famiglia e Francesca per avermi supportato in questi anni di studio e per l’affetto con cui incoraggiano e sostengono le mie scelte più importanti.

Antonio Di Pippo

INDICE

	pag.
<i>INTRODUZIONE</i>	6
 <i>CAPITOLO 1 – L'ANALISI ESPLORATIVA DEI DATI</i>	
1.1 Il Sistema Informativo Aziendale	9
1.2 Dai Database, ai Data Warehouse...	11
1.2.1 L'architettura dei sistemi di Business Intelligence	14
1.3 ...al Data Mining	21
 <i>CAPITOLO 2 – APPLICAZIONI STATISTICHE IN AZIENDA</i>	
2.1 A cosa servono i dati raccolti	28
2.2 Tecniche di indagine statistica	33

2.3	Eventuali ambiti applicativi	35
2.3.1	Valutazione degli effetti delle promozioni di vendita	36
2.3.2	Metodi associativi per la Market Basket Analysis	42
2.3.3	Sondaggi di opinioni e ricerche di mercato per la Customer Satisfaction	48
2.3.4	La gestione della qualità: il Total Quality Management	55
2.3.5	Misura dell'audience e dell'efficacia della pubblicità	60
2.3.6	Analisi delle sequenze di visita ad un sito di e-commerce	67
2.4	L'obiettivo: il CRM	72

CAPITOLO – 3 IL CASO “MasterDEA”

3.1	I servizi di formazione universitaria	79
3.2	Il master: una scelta delicata in un mercato del lavoro incerto	81
3.3	I MasterDEA	84
3.4	Il mercato della formazione	87

3.5 Il caso	89
3.5.1 Le informazioni	90
3.6 L'analisi statistica dei dati	95
3.6.1 L'analisi "prima – dopo"	97
3.6.2 Analisi di connessione tra i canali informativi	100
3.6.3 L'ANOVA a due criteri di classificazione: canale informativo – master	104
3.6.4 Le differenze tra i master di primo e di secondo livello	110
3.6.5 L'analisi di correlazione	116
3.6.6 L'analisi finanziaria	120

CAPITOLO 4 – CONCLUSIONI

4.1 I risultati dell'analisi	125
4.2 Considerazioni	130

<i>BIBLIOGRAFIA</i>	133
---------------------	-----

INTRODUZIONE

Nell'attuale società dell'informazione, la quantità e la disponibilità di dati è cresciuta a dismisura relativamente alle informazioni interne all'azienda, e a quelle acquisite dall'esterno. Competere sull'informazione è, dunque, diventato un dovere e ottenere le informazioni migliori consente vantaggi competitivi notevoli. Oggi gli strumenti quantitativi alla base delle decisioni di management assumono sempre più importanza grazie anche allo sviluppo dei software che rendono più avvicinabili talune metodologie anche a coloro che hanno competenze limitate in materia di elaborazione dati.

All'interno delle aziende, mentre cresce l'attenzione per la gestione dell'informazione e, di conseguenza la domanda di statistiche economiche generali, l'impiego di metodi di analisi quantitativa e dei software correlati resta ancora al di sotto delle sue effettive potenzialità. Alla radice di questo stato di cose vi sono vari motivi: da un lato, la scarsa sensibilità statistica di una parte dei manager poco inclini a ragionare e a decidere in base a modelli concettuali di tipo probabilistico e statistico; dall'altro una modellistica statistica che non riesce a cogliere tutti gli aspetti essenziali delle situazioni reali e la difficoltà di misurare con approssimazione soddisfacente gli aggregati economici e finanziari necessari per elaborare modelli. Solo nell'area del *marketing* e delle analisi di mercato la situazione è un po' diversa. L'instabilità e la competitività dei mercati inducono le aziende, che adottano strategie commerciali, a seguire con attenzione i comportamenti di acquisto dei consumatori e le loro reazioni alle iniziative di *marketing-mix* tramite indagini

generalmente impostate sotto forma di sondaggi, secondo una metodologia che si avvale segnatamente dell'apparato strumentale della statistica e della statistica economica.

In azienda le decisioni razionalmente fondate, da intendersi nel senso di azioni consapevoli e intenzionali, devono ovviamente basarsi su informazioni soprattutto di tipo statistico, che risultino al tempo stesso adeguate e di buona qualità. Strumenti fondamentali per produrre, selezionare ed elaborare informazioni sono in maniera evidente i metodi statistici e le tecnologie informatiche. La statistica aziendale è configurabile come l'insieme a largo spettro dei metodi statistici utilizzabili per l'organizzazione e l'analisi dei dati aziendali di supporto alle decisioni manageriali.

Il percorso logico seguito per lo sviluppo di questo lavoro parte dalla volontà di ampliare le conoscenze statistiche al fine di trarne indicazioni per una valida applicazione in contesti, quali quelli aziendali.

Il lavoro si compone principalmente di due parti. La prima di carattere prevalentemente descrittivo e metodologico (capitoli 1 e 2), la seconda di carattere applicativo (capitolo 3).

Nel primo capitolo, "*L'analisi esplorativa dei dati*", si è tentato di descrivere in modo sistematico la struttura e il funzionamento dei sistemi informativi aziendali ed in particolare l'attività di *Data Mining* (estrazione dati), intesa come processo di analisi di basi di dati orientato a produrre dei risultati utili nei processi decisionali. In questo capitolo vengono illustrati alcuni aspetti preliminari relativi all'organizzazione dei dati per predisporli al caricamento nei *Data Warehouse* (insieme di basi di dati) e al loro successivo utilizzo per produrre report di sintesi utili ai livelli direzionali.

Il secondo capitolo, "*Applicazioni statistiche in azienda*", presenta una selezione di problemi di natura decisionale e gestionale che in azienda sono

inevitabilmente trattati in condizioni di incertezza e di rischio, e che trovano quindi nei metodi statistici un supporto indispensabile e scientificamente rigoroso per la loro soluzione. I temi affrontati riguardano la valutazione degli effetti delle promozioni di vendita, i metodi associativi per la *Market Basket Analysis*, la gestione statistica della soddisfazione della clientela (*Customer Satisfaction*) e della qualità totale (*Total Quality Management*), il controllo dell'efficacia della pubblicità e l'analisi statistica delle sequenze di visita ad un sito di *e-commerce*.

Nel terzo capitolo si è analizzata, in tutti i suoi aspetti, la strategia pubblicitaria del Dipartimento di Economia Aziendale dell'Università di Pisa messa in atto per promuovere l'offerta formativa post-laurea "*MasterDEA*". Dopo una breve digressione sul mercato della formazione e sulla struttura organizzativa dei master, si è cercato di dare risposta ad alcuni quesiti fondamentali:

- i mezzi di comunicazione utilizzati sono tutti validi?
- esiste un mezzo più efficace degli altri?
- sussiste una connessione tra i comportamenti di visita on-line e il numero di soggetti che presenta domanda di iscrizione?

Solo rispondendo a queste domande, sarà possibile stabilire e individuare le possibili aree di miglioramento della comunicazione pubblicitaria da mettere in atto in un settore, quello della formazione post-laurea, che negli ultimi anni sta sperimentando l'inasprirsi della concorrenza.

Il quarto ed ultimo capitolo è un compendio dei risultati conseguiti con le analisi statistiche sulla pubblicità "*MasterDEA*" alla luce della logica che ha guidato tutto il presente lavoro: palesare ancora una volta come la statistica sia un valido supporto all'attività manageriale.

1

L'analisi esplorativa dei dati

1.1

Il Sistema Informativo Aziendale

Con il termine *Sistema Informativo Aziendale* (SIA) si è soliti indicare quell'insieme di flussi di informazioni all'interno di un'organizzazione.

Così come definiti in letteratura¹, i SIA sono composti non semplicemente dalle soluzioni tecnologiche hardware e software o dalle basi dati che li realizzano, ma anche dalle persone che li utilizzano o li creano, dalle norme, dalle procedure organizzative e dai meccanismi operativi che li fanno funzionare, evolvere e diffondere in azienda.

A partire dalla seconda metà degli anni Sessanta, il progresso e una serie di sviluppi tecnologici hanno aumentato il fabbisogno informativo delle imprese. Fra essi lo sviluppo di tecniche per il controllo della gestione, che richiedono l'elaborazione di un'enorme quantità di dati e il processo di

¹ Pasini M., *I sistemi informativi direzionali. Le tecnologie dell'informazione a supporto dei processi manageriali d'azienda*, EGEA, Milano, 2004.

internazionalizzazione delle organizzazioni e la generale terziarizzazione dell'intero tessuto economico, che aumentano il fabbisogno di dati per coordinare e controllare le attività aziendali.

L'emergere di queste tendenze e la crescente disponibilità di dati nell'attuale società dell'informazione ha, dunque, evidenziato la necessità di disporre di strumenti adeguati per la loro analisi.

La statistica applicata e il *Data Mining* si propongono come mezzi privilegiati per estrarre informazioni da questi dati e rapidamente si stanno diffondendo in molte organizzazioni aziendali come importanti strumenti di business intelligence, ovvero di tecnologia dell'informazione per il supporto alle decisioni.

Pur nella loro grandissima varietà i sistemi informativi aziendali (SIA) hanno molte caratteristiche in comune per quel che concerne le informazioni che:

- riguardano il funzionamento interno dell'impresa;
- sono strutturate secondo formati predeterminati;
- sono correlate;
- forniscono ai livelli gerarchici e alle varie unità organizzative dell'impresa una rete informativa su cui si basano le operazioni e le decisioni.

In particolare, all'interno di un'azienda è fondamentale, affinché sistema informativo possa integrarsi al meglio, che le persone che ne fanno parte possiedano le seguenti caratteristiche:

- entusiasmo e motivazione;
- capacità di svolgere il loro lavoro in modo integrato con gli altri;
- capacità di accettare il cambiamento nel proprio modo di lavorare.

Da un lato, quindi, l'azienda stessa dovrebbe essere in grado di esplicitare le proprie necessità al sistema informativo, dall'altro i sistemi informativi dovrebbero essere non soltanto un supporto tecnico, ma possedere una profonda conoscenza del sistema azienda in modo da:

- fornire ai gestori informazioni qualitative, oltre che quantitative, al momento giusto e nella forma giusta;
- poggiare su tecnologie che consentano la necessaria flessibilità;
- operare una gestione anticipata (pianificazione degli investimenti);
- operare una gestione con visione strategica e consapevole del rischio.

1.2

Dai Database, ai Data Warehouse...

Ogni anno la quantità totale di dati disponibili nel mondo, in formato digitale, aumenta e solo in minima parte viene utilizzata da istituzioni, organizzazioni ed aziende; si pensi, ad esempio, ad internet che produce giornalmente grandi quantità di dati che aspettano di essere trasformati in informazioni tramite sistemi di supporto alle decisioni e analisi statistiche applicate.

Da qualche anno il mercato informatico ha recepito questa esigenza sviluppando sistemi di supporto alle decisioni in grado di fornire, in tempo reale, informazioni, rapporti e consentire analisi di varia natura, tra le quali la *What If Analysis*, che permette previsioni basate su ipotesi sui dati futuri; l'*On Line Analytical Processing* (OLAP), che mette a disposizione del manager un ambiente di dati multidimensionale, nel quale può eseguire ricerche aggregando i dati in suo possesso; ed il *Data Mining*, che applica tecniche di intelligenza artificiale agli archivi aziendali alla ricerca di quei dati che non sono visibili in un primo istante perché nascosti in un'enorme quantità di dati simili.

L'infrastruttura alla base dei sistemi di supporto alle decisioni è il *Data Warehouse* che in letteratura² è definito come l'anello di collegamento tra i dati, le applicazioni e i sistemi informativi di tipo operativo e transazionale e i sistemi informativi manageriali di supporto alle attività di controllo e di decisione.

A differenza di quanto avviene nelle applicazioni tipiche dei database, dove i moduli di inserimento ed estrazione sono già predefiniti ed è quindi praticamente impossibile combinare i dati per analisi statistiche rilevanti ai fini aziendali, il data warehouse è un "magazzino" nel quale sono registrati i dati provenienti da molte fonti correlate o meno tra loro. Tale "magazzino" è indipendente dagli altri archivi del sistema e può essere interrogato liberamente, ossia non esiste uno schema predefinito di domande, ma è l'utente che di volta in volta costruisce dinamicamente la propria interrogazione a seconda delle esigenze di analisi.

Il data warehouse è, dunque, un'evoluzione dei più semplici sistemi di database e il suo grande vantaggio sta nella possibilità di sfruttare, per il supporto decisionale a tutti i livelli, dati e informazioni che già si trovano all'interno dell'organizzazione. Esso risolve problemi di informazione frammentata in numerose tabelle, di elaborazioni spesso molto lunghe e di volatilità dell'informazione dovuta alla continua modifica dei dati gestionali che impedisce di ripetere l'analisi anche in tempi brevi ottenendo i medesimi risultati.

Una valida struttura di data warehouse, quindi, dovrebbe prevedere le seguenti componenti:

- un archivio centralizzato, che costituisce il "magazzino" dei dati;
- una struttura di dati che descriva cosa e dove è disponibile nel data warehouse;

² Albano A., *Costruire sistemi per basi di dati*, Addison Wesley Longman Italia, Milano, 2001.

- una serie di dati specifici e tematici, specializzati a seconda degli obiettivi di business intelligence, rapidamente accessibili e convertibili in strutture statistiche.

L'integrazione è la più importante delle caratteristiche di un data warehouse e nasce dalla necessità di dare coerenza ai dati provenienti da diverse applicazioni progettate per scopi diversi. I data warehouse si basano sull'utilizzo misto di svariati modelli (relazionale, gerarchico, multidimensionale) e su svariate tipologie di archiviazione (tabelle, report, grafici, testi, immagini) nell'ottica di consentire la massima flessibilità di scelta nell'esplorazione e nello sfruttamento delle informazioni. Il vantaggio competitivo ottenibile attraverso questo tipo di sistema per il supporto alle decisioni sta nella possibilità di estrarre, da una massa di dati poco gestibile, informazioni strategicamente utili ed allineate con il modello di attività del decisore.

All'interno del sistema informativo aziendale (SIA), è necessario che il data warehouse assuma tre ruoli sostanziali:

1. ruolo di "integratore" tra sistemi informativi e basi di dati operativi differenti e stratificate nel tempo, interne ed esterne, correnti e storiche;
2. ruolo di "disaccoppiatore" tra i sistemi informativi operativi, che automatizzano i processi operativi aziendali di tipo routinario e ripetitivo, e i SIA, che supportano attività e processi di controllo e di decisione manageriale;
3. ruolo di "consolidatore" nelle strutture di gruppo, nelle organizzazioni "multidivisionali", "multibusiness" o "multigeografiche"; infatti il data warehouse può permettere a livello centrale (sede centrale o holding operativa) di omogeneizzare e analizzare correttamente fenomeni gestionali complessi generati da politiche e procedure gestionali diversificate (per prodotto o per area geografica).

Esistono sostanzialmente due diversi tipi di approccio alla creazione di un data warehouse.

Il primo si basa sulla creazione di un unico archivio centralizzato che raccoglie tutte le informazioni aziendali e le integra con quelle provenienti dall'esterno.

Il secondo, invece, nasce dall'unione di diversi database inizialmente non collegati tra loro, che evolvono verso una struttura perfettamente interconnessa.

L'approccio centralizzato permette un costante controllo da parte degli amministratori del sistema della qualità dei dati che vengono immessi. D'altra parte, però, richiede un'attenta progettazione che ne permetta una futura espansione, sia per quanto riguarda la capacità di acquisire nuovi dati, sia per la possibilità di collegarsi ad altri data warehouse.

Il secondo approccio invece, è relativamente più semplice nelle fasi di implementazione iniziale e per questo risulta essere la soluzione più diffusa attualmente. I problemi iniziano a sorgere quando i vari database vengono collegati tra loro: bisogna compiere un notevole sforzo di definizione per ottenere un sufficiente livello di uniformità.

1.2.1

L'architettura dei sistemi di Business Intelligence

Dal punto di vista dell'architettura tecnologica si possono identificare tre livelli base (fig. 1)³:

- sistemi alimentanti, ovvero i sistemi che contengono i dati elementari;

³ Pasini M., *I sistemi informativi direzionali. Le tecnologie dell'informazione a supporto dei processi manageriali d'azienda*, op. cit.

- sistemi di data warehouse, ovvero i sistemi per l'integrazione e l'archiviazione dei dati semilavorati;
- funzionalità di business intelligence, ovvero i sistemi per l'accesso ai dati e la produzione di informazioni "finite".

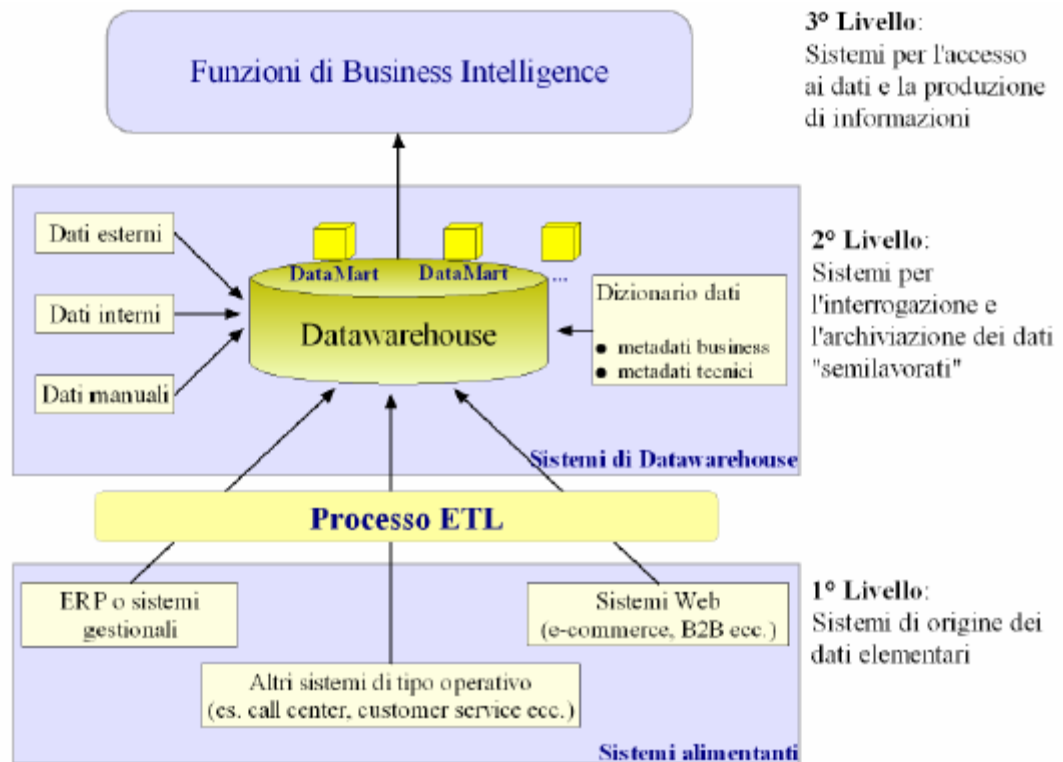


Figura 1: L'architettura dei 3livelli dei sistemi di *Business Intelligence*

I sistemi alimentanti

I sistemi alimentanti o sistemi di origine dei dati elementari, sono costituiti dai sistemi gestionali e amministrativo-contabili di tipo ERP o tradizionali, dai sistemi che interfacciano il mercato (sistemi di CRM), dai sistemi Web (portali, e-commerce, siti informativi) e da tutti gli altri sistemi informativi formali di tipo operativo e/o transazionali (file di testo, fogli di calcolo).

I *sistemi ERP* (Enterprise Resource Planning) o sistemi informativi integrati sono dei pacchetti software le cui applicazioni coprono in maniera integrata le esigenze informative delle varie funzioni di un'azienda (ad es. amministrazione, produzione, distribuzione, marketing, personale ecc.) e permettono di pianificare, gestire e organizzare in modo unitario le risorse dell'azienda. Essi sono integrati, sia sotto il profilo dell'architettura informatica, sia sotto quello della progettazione logica, nel senso che permettono soluzioni di integrazione di tutti i processi gestionali e amministrativi.

I *sistemi di CRM* (Customer Relationship Management) permettono di:

- automatizzare e ottimizzare le attività operative di marketing, di vendita e di customer service su canali off-line e on-line di interazione con il mercato;
- integrare le molteplici fonti di dati sul cliente, modellarli, organizzarli e analizzarli al fine di creare nuova conoscenza aziendale sui clienti e sulle loro relazioni nel ciclo di vita e supportare più efficaci decisioni (di marketing, di vendita e di servizio) che impattano direttamente sul valore dei clienti attuali o prospectivi.

I dati elementari contenuti in questi sistemi vengono estratti, puliti e trasformati e infine caricati nel data warehouse. Quest'ultima fase si pone logicamente tra il primo ed il secondo livello ed è meglio conosciuta come processo di ETL (Extraction, Transformation, Loading).

Processo ETL

Con l'avvento di Internet, di nuove tecnologie sempre più avanzate e l'esigenza da parte di aziende, università e amministrazioni pubbliche di

operare in un contesto competitivo sempre più ricco di informazioni ma soprattutto di dati spesso provenienti da fonti eterogenee, cresce sempre più il bisogno di integrare completamente le diverse applicazioni al fine di supportare le attività decisionali. Il problema quindi è quello di riuscire ad integrare le informazioni provenienti dalle fonti più disparate (sistemi ERP, mainframe, basi di dati relazionali e file di applicativi) e di muoverli attraverso i diversi ambienti dell'azienda o istituzione o, eventualmente, all'esterno di essa consentendo di indirizzare le scelte del management.

È proprio in questo contesto che trovano spazio le applicazioni *ETL* che vengono impiegate per rendere disponibili i dati raccolti in azienda ai soggetti incaricati di assumere le decisioni, nella forma e secondo le tempistiche più idonee a supportare il processo decisionale.

Per riuscire a sfruttare pienamente i dati operativi, l'infrastruttura tecnica preposta al sostegno dei processi decisionali deve essere in grado di raccogliere, consolidare, trasformare e trasferire i dati, predisponendoli al meglio per la successiva fase analitica.

Il processo ETL consta di tre fasi principali (fig. 2)⁴:

- estrazione
- pulizia e trasformazione
- caricamento

⁴ Pasini M., *I sistemi informativi direzionali. Le tecnologie dell'informazione a supporto dei processi manageriali d'azienda*, op. cit.

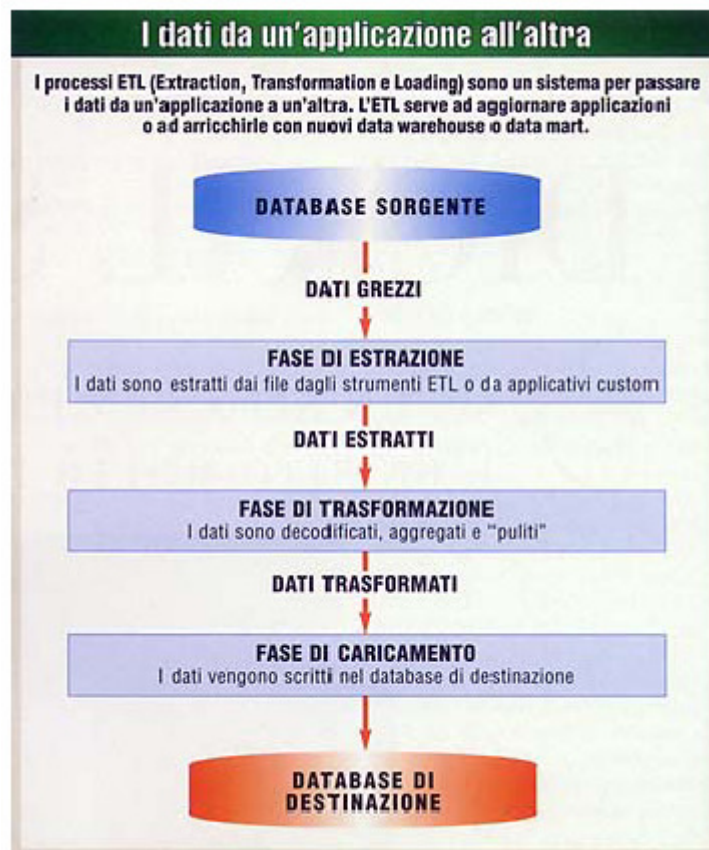


Figura 2: Fasi di un processo di ETL

Vediamo nel dettaglio le tre fasi sopra citate.

Estrazione

Lo scopo del processo di estrazione dei dati consiste nel convertire tutti i dati di origine in un formato comune e consistente per prepararli per il caricamento nel data warehouse.

Questa fase implica la predisposizione di procedure in grado di leggere i record contenuti nei database o file sorgente e di predisporli per la successiva fase di trasformazione. I dati vengono inizialmente estratti durante la creazione del data warehouse, quindi vengono nuovamente estratti periodicamente durante gli aggiornamenti. L'estrazione dei dati può essere un'operazione semplice, se

i dati di origine si trovano in un unico database, oppure molto complessa, se i dati di origine si trovano in più sistemi attivi ed eterogenei.

Tre sono le alternative che è possibile seguire nella fase di estrazione dei dati: gli sviluppatori potranno scrivere programmi ad hoc, scegliere in alternativa di affidarsi agli strumenti ETL presenti sul mercato (ad es. DTS di SQL Server 2000) o utilizzare entrambe le soluzioni.

Pulizia e trasformazione

Questa fase consiste nel pulire, trasformare e integrare i dati. Dopo che questi sono stati raccolti, infatti, le procedure di trasformazione provvedono a prepararli per la loro nuova collocazione.

Con l'operazione di pulitura si determina la qualità dei dati che verranno caricati nel data warehouse correggendo gli errori e le inconsistenze presenti nei dati elementari (ad es. dati duplicati, dati mancanti, valori errati o nulli). Sono, infatti, numerosi gli errori che sono difficilmente evitabili se non si prevedono procedure di controllo dei valori inseriti; altrettanto comune è riscontrare differenze di codifica dei dati nello stesso campo (ad es. utilizzare abbreviazioni differenti per indicare lo stesso comune o la stessa nazione).

Spesso la correzione delle inconsistenze tra dati provenienti da origini eterogenee non viene eseguita fino a quando i dati non sono stati raccolti in un database comune: è solo a questo punto, infatti, che risulta più semplice individuare e correggere gli errori di integrità dei dati.

Per supportare le operazioni di ripulitura e di trasformazione e per preparare i dati per il caricamento nel data warehouse è spesso necessario creare tabelle e altri database: l'obiettivo è quello di automatizzare al massimo il processo e di gestire e aggiornare gli strumenti sviluppati.

Ovviamente la struttura specifica necessaria alla preparazione dei dati dipende delle origini dati, dal grado di trasformazione necessario per organizzare i dati e dalla consistenza dei dati di arrivo. In alcuni casi, infatti, è opportuno modificare i sistemi di origine dei dati in modo da eliminare la causa del problema. Si tratta di operazioni di formattazione da compiere prima di caricare i dati nel data warehouse.

Caricamento

Rappresenta l'ultima fase del processo ETL e consiste nel caricare le informazioni nel data warehouse di destinazione. Durante questa fase il programmatore si trova a dover scegliere l'impostazione del tipo di caricamento, ovvero se il caricamento nella base di dati deve avvenire in forma periodica oppure in forma continuativa. Questa ultima opzione risulta, sovente, costosa poiché richiede reti dedicate ad alta velocità.

Nel data warehouse vengono caricati dati storici, dati correnti, spesso dati previsionali riferiti all'immediato futuro, e dati esterni. Questi dati devono essere "semanticamente corretti" (non devono cioè esistere dati diversi identificati con lo stesso nome) e devono essere rilevati e calcolati con criteri omogenei nel tempo (cioè invariati nel tempo per poter confrontare i dati passati con i dati correnti) e nello spazio (cioè uguali nelle diverse funzioni, divisioni, unità operative, magazzini o filiali dell'azienda). A tal fine nel data warehouse deve esserci un catalogo dati che riassume il significato preciso di ogni dato, le sue modalità di calcolo e la sua certificazione in termini di fonte di origine e allineamento rispetto a essa, di proprietà del dato e di omogeneità (ad es. unità di misura).

1.3

... al Data Mining

Da diversi anni ormai, stanno diventando sempre più familiari termini quali *Data Mining* e *KDD*, spesso confusi l'uno con l'altro, ma cosa sono in realtà?

Di fronte a una situazione di abbondanza di dati e scarsità di informazioni, ogni individuo sia esso personale, professionale o aziendale, ha l'esigenza di avere a disposizione tecniche e strumenti capaci di analizzare in modo automatico e intelligente queste grandi "miniére" di dati, al fine di "estrarre" la conoscenza ivi nascosta.

Queste tecniche e strumenti sono rappresentate dall'emergente campo del *KDD* (*Knowledge Discovery in Databases*). Il termine *KDD* viene utilizzato per descrivere l'intero processo di estrazione della conoscenza da un database, dall'individuazione degli obiettivi di business iniziali fino all'applicazione delle regole decisionali. Tecnicamente *KDD* è il processo per identificare nei dati, pattern (struttura/modello) con caratteristiche di validità, novità, utilità potenziale e facilità di comprensione.

Il *Data Mining* è proprio la componente principale del processo *KDD*, è la fase di applicazione di uno specifico algoritmo per l'individuazione dei "pattern" nascosti dei dati.

Una definizione più completa di data mining è pertanto la seguente⁵:

"Per Data Mining si intende il processo di selezione, esplorazione e modellazione di grandi masse di dati, al fine di scoprire regolarità o relazioni non note a priori, e allo scopo di ottenere un risultato chiaro utile al proprietario del database".

⁵ Giudici P., *Data mining: metodi statistici per le applicazioni aziendali*, Mc Graw-Hill, Milano, 2001.

In altre parole, con questo termine si può intendere l'applicazione di una o più tecniche che consentono l'esplorazione di grandi quantità di dati e che danno risposte a domande di carattere generico senza richiedere la formulazione di ipotesi a priori. È un algoritmo che si fa carico di mettere in evidenza le variabili e le caratteristiche che si presentano ripetutamente, senza richiedere all'utente la formulazione di un'ipotesi a priori: si tratta di un approccio esplorativo e non verificativo. L'obiettivo è quello di individuare le informazioni più significative e di renderle disponibili e direttamente utilizzabili nell'ambito del *decision making*.

Il processo di *KDD* è un processo interattivo e iterativo suddiviso in diverse fasi (fig. 3)⁶:

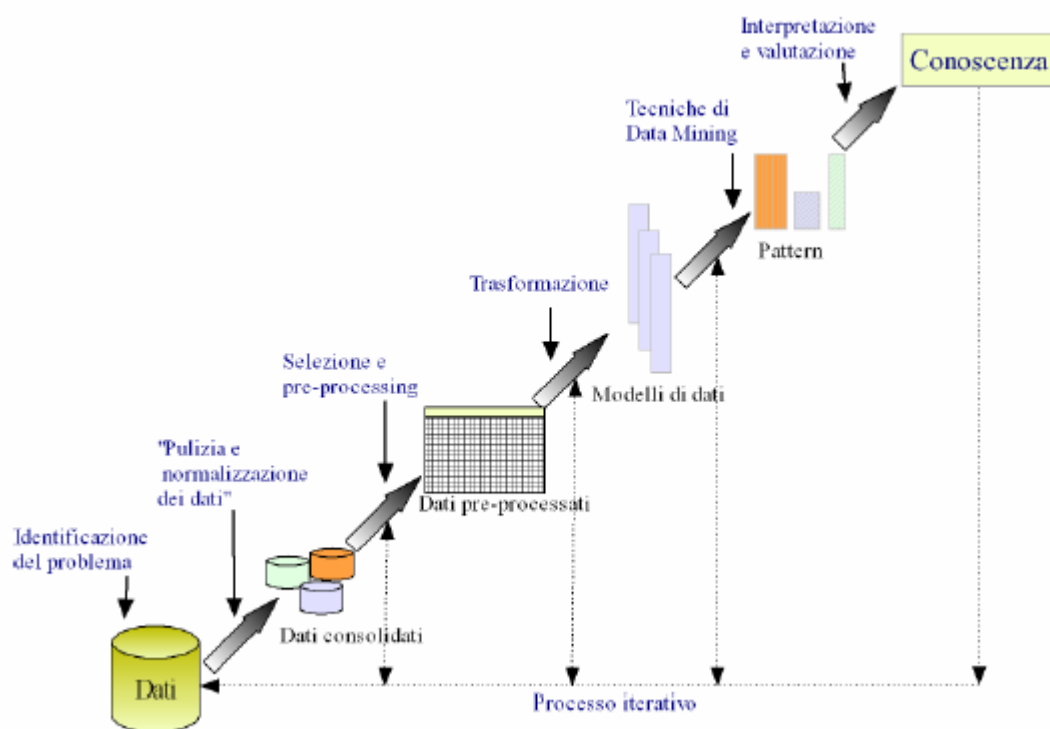


Figura 3: Processo di *KDD*.

⁶ Del Ciello N., Dulli S., Saccaridi A., *Metodi di data mining per il customer relationship management*, Franco Angeli, Milano, 2000.

Identificazione del problema

Si identifica il problema, tenendo conto della conoscenza già acquisita in precedenza e degli obiettivi che si vogliono perseguire. È evidente che il flusso di lavoro, la scelta dei dati e la comprensione dei risultati conseguiti, dipendono dagli obiettivi che si intendono raggiungere.

Sarà pertanto necessario tradurre tali obiettivi in obiettivi di analisi, eliminando dubbi o incertezze su cosa si intende perseguire e sulle modalità che verranno impiegate.

La definizione di obiettivi precisi e controllabili è anche necessaria al fine di una corretta definizione del problema di data mining e di un disegno preliminare di analisi.

Pulizia e normalizzazione dei dati

In un'azienda, le fonti da cui attingere i dati interessanti per l'analisi sono molteplici.

È dunque necessaria una procedura che trasformi i dati rilevati in dati elaborabili. A tal fine vengono effettuate diverse operazioni tra cui la creazione di tabelle, la selezione di record, la trasformazione e ricodifica di variabili e il calcolo di nuovi indicatori.

Chiaramente la presenza di un data warehouse facilita la pulizia e la normalizzazione dei dati attraverso, ad es. l'eliminazione dei dati rumorosi (noise) e dei valori estremi (outlier), la gestione dei campi vuoti (missing values field).

Selezione dei dati appropriati per l'analisi

In questa fase si deve decidere quali e quanti dati tra quelli disponibili trattare nell'analisi. Questa operazione è molto delicata perché dipende fortemente dallo stato e dalla quantità di dati a disposizione e perché non esistono regole precise che guidano l'analista nel fare la scelta migliore. Per quanto riguarda la numerosità l'insieme dei dati deve contenere tutti dati necessari per descrivere i vari aspetti del fenomeno, anche quelli più marginali. Inizialmente è bene non escludere a priori nessuna variabile sulla base della conoscenza intuitiva e aleatoria del business ma aspettare che siano i dati stessi a rivelare ciò che è importante o meno. Riguardo alla storicità, poiché il processo si basa sull'apprendimento dalle esperienze passate, è necessario avere i dati sui comportamenti passati dei clienti e dell'azienda.

Estrapolazione

Si individuano le caratteristiche salienti per rappresentare il fenomeno che si sta analizzando in funzione dell'obiettivo definito, tendendo a ridurre il numero delle variabili prese in considerazione.

Prima di iniziare a costruire il modello è quindi indispensabile conoscere i propri dati e il contesto di business. Le analisi di tipo multidimensionale sono di grande aiuto per osservare i dati da diversi punti di vista e offrono sintesi interessanti per approfondire la conoscenza del business. E', inoltre, molto utile che per ogni variabile vengano calcolate la frequenza dei valori, la media, la mediana e vengano eseguite tutta una serie di operazioni di controllo e visualizzazione dei dati per assicurarsi che i dati siano trattabili.

Creazione di un modello di analisi

Un modello è una rappresentazione concettuale che descrive e spiega le regole con cui la realtà funziona e che può essere usato per fare delle inferenze su come andranno le cose nel mondo reale.

Nella fase di creazione di un modello è necessario definire il cosiddetto “*Data Mining task*”, cioè il tipo di analisi da effettuare sui dati (classificazione, previsione, clustering, regole associative); scegliere le tecniche di data mining da impiegare per ricercare i pattern nei dati, in funzione del criterio generale alla base del processo di *KDD* (ad esempio, l’analista potrebbe essere maggiormente interessato alla comprensione del modello rispetto alle capacità di previsione dello stesso); ed infine eseguire il data mining, cioè compiere la ricerca dei pattern di interesse.

La scelta di quale tecnica o combinazioni di tecniche utilizzare dipende dal tipo di problema e dal tipo di dati disponibili.

Valutazione del modello applicato

Per valutare il modello si prendono in considerazione alcune caratteristiche tecniche come l’accuratezza, il livello di sicurezza sull’attendibilità dei risultati, ma anche considerazioni come la capacità espressiva, la semplicità e la comprensibilità del modello. I parametri usati per misurare l’accuratezza di un modello o il suo grado di confidenza variano a seconda del tipo di modello implementato, così come varia l’importanza attribuita alla semplicità e comprensibilità del modello e dei suoi risultati a seconda degli obiettivi iniziali.

Interpretazione dei risultati

Interpretare i risultati dell'analisi del data mining è un momento critico del processo di analisi, sia perché i risultati sono di difficile lettura e comprensione sia perché l'output del data mining non dà informazioni sempre utili o significative per il business. In questa fase l'analista deve condividere e illustrare i risultati al manager, al fine di giungere all'individuazione di informazioni davvero rilevanti ai fini dell'analisi.

Implementazione delle regole nei processi decisionali

L'integrazione dei risultati nei processi decisionali aziendali è il passo finale del processo e coinvolge i manager.

In questa fase si consolida e si formalizza la conoscenza acquisita (realizzazione/integrazione di un sistema applicativo, redazione di documentazione, presentazione alle parti interessate, ecc...).

Nonostante il campo di applicazione delle tecniche di data mining sia in continuo mutamento, alcune strategie fondamentali sono rimaste invariate. Per tecnica si intende uno o più algoritmi con una struttura conoscitiva che applicano una specifica strategia a un insieme di dati; per strategia, invece, si deve intendere uno schema (linee guida) di approccio alla soluzione di un problema.

Esistono due tipologie di strategie di data mining:

- strategie supervisionate, in cui vengono costruiti dei modelli sulla base di attributi in input per predire o classificare i valori degli attributi di output;

- strategie non supervisionate, in cui vengono costruiti dei modelli sulla base di attributi di input, ma non esiste alcun attributo di output da predire.

Gli attributi di output vengono definiti variabili dipendenti in quanto i loro valori dipendono dai valori di uno o più attributi di input (questi ultimi vengono definiti variabili indipendenti).

Le tecniche attraverso cui il data mining viene realizzato si basano su metodologie statistiche e matematiche e si distinguono in: alberi decisionali (o decision tree); reti neurali e regole di associazione per le strategie di apprendimento supervisionato; e cluster analysis per le strategie di apprendimento non supervisionato.

Il data mining trova diverse applicazioni nell'ambito dell'economia e della finanza, nella scoperta di frodi bancarie, in applicazioni scientifiche, nella cura della salute e in sport e giochi (per esempio l'industria dei giochi ha incorporato modelli storici dei trend dei giocatori d'azzardo per individuare quanto un singolo giocatore sarà disposto a spendere durante una sua visita al casinò).

Uno degli ambiti in cui l'applicazione del data mining si rivela particolarmente proficua è il Marketing. La strategia applicata in questo caso è quella della *Market Basket Analysis* (MBA), il cui scopo è quello di ricercare relazioni interessanti tra i prodotti delle vendite al dettaglio al fine di sviluppare successivamente strategie di marketing incrociato.

In generale il processo di analisi con il data mining apporta una grande quantità di informazioni, dalla fase iniziale di preparazione dei dati a quella finale di interpretazione, e anche nel caso in cui non emergano aspetti particolarmente sorprendenti, aiuta i manager e l'analista ad osservare e ragionare sui dati da un punto di vista diverso.

2

Applicazioni statistiche in azienda

2.1

A cosa servono i dati raccolti

Il data mining è stato definito come un processo, all'interno del quale si utilizzano una o più tecniche per ottenere dai dati indicazioni a supporto dell'attività di business. Si tratta di un processo che estrae conoscenza attraverso l'individuazione di regole e associazioni tra i dati non note e non visibili ad occhio nudo.

Fare data mining significa seguire un processo che va dalla traduzione della problematica di business in un obiettivo di analisi quantitativa, all'implementazione di regole decisionali misurabili. La scelta di quale tecnica utilizzare nella fase di analisi dipende, essenzialmente, dal tipo di problema oggetto di studio e dal tipo di dati disponibili per l'analisi.

Il data mining è qualcosa di molto più complesso di una tecnica di analisi. È difficile da implementare, è difficile da gestire ma può portare a risultati molto importanti con un impatto (positivo) sul business molto forte.

Per capire se e perché un'azienda dovrebbe avere bisogno di un sistema di data mining è necessario capire che tipo di risposte può dare un sistema di data mining. Dal processo di data mining dovranno, infatti, emergere con chiarezza:

1. gli obiettivi di business;
2. la definizione dei relativi problemi di data mining;
3. le risorse (tecnologiche, umane, finanziarie) disponibili e non;
4. lo stato dei dati.

Il data mining applicato nel contesto del marketing è usato per migliorare il business sia dal punto di vista del costo che dal punto di vista della redditività; il data mining è, infatti, usato per diminuire i costi delle campagne marketing evitando di contattare clienti che probabilmente risponderebbero negativamente, ma è anche usato per incrementare la redditività attraverso la calibratura dell'offerta su quelle classi di clienti con caratteristiche che indicano che sono pronte per essere spinte verso un comportamento più profittevole.

Una volta capito se un sistema di data mining può fare al caso del proprio business, bisogna capire e valutare se l'azienda dispone delle risorse e delle capacità necessarie ad implementarlo e se il costo e i rischi di adozione di un tale sistema siano pienamente ripagati dai benefici che verrebbero generati.

Per quanto riguarda lo stato dei dati, invece, c'è da chiedersi: *quanti dati servono per fare data mining?*

In generale è necessario che i dati siano numericamente e qualitativamente sufficienti a rappresentare, in tutti i suoi aspetti, la realtà di business che si vuole analizzare; avere milioni di dati scorretti o, dove mancano le informazioni critiche che si vogliono indagare, equivale a non avere dati. I dati

devono contenere, inoltre, tutte le informazioni necessarie ai fini dell'analisi e degli obiettivi di data mining prefissati. Sarebbe, quindi, utile avere:

- dati che riguardano la vita del cliente: di tipo anagrafico e demografico, variabili socio-economiche, comportamentali, sulle località di residenza.
- dati che riguardano il settore e il business (*prospect* anonimi): provenienti da indagini di mercato, ricerche, ecc.
- dati che riguardano l'azienda: budget, risultati economici, personale, organizzazione, ecc.
- dati provenienti dai punti di contatto tra l'azienda e il cliente: call center, posta elettronica e tradizionale, sito web, negozio.
- dati sulle campagne marketing: descrizione delle campagne, risposte o non risposte dei contatti.
- dati di ritorno dalla *customer satisfaction audit*.

In una qualsiasi organizzazione, aziendale e non, il data mining fornisce risposte valide a problematiche di natura strategica più efficacemente di qualunque altro sistema di reporting. Fondamentalmente tali problematiche possono essere raggruppate nelle seguenti categorie.

Descrizione. Il data mining può essere semplicemente utilizzato per descrivere relazioni all'interno di un database, così da aumentare la conoscenza del fenomeno oggetto di studio. L'obiettivo, in generale, è quello di identificare relazioni complesse e ad alta dimensionalità al fine di scoprire l'informazione nascosta.

Classificazione. La classificazione rappresenta la più comune attività del data mining e consiste nell'identificare schemi o insiemi di caratteristiche che definiscono una classe e nell'esaminare gli attributi di un oggetto per

assegnarlo a una classe predefinita. Gli elementi da classificare sono generalmente rappresentati da record contenuti in un database.

Stima. Si utilizza l'attività di stima per dare un valore ad un insieme di output non identificati. Per stimare la probabilità di risposta positiva a una certa offerta, per esempio, si potrebbero selezionare i clienti in base al possesso di un determinato prodotto o a determinati comportamenti. Anche alcune attività di classificazione possono essere considerate di stima e sono, comunque, estremamente collegate: la previsione della perdita di un segmento di clientela, per esempio, può essere considerata sia come classificazione dei clienti in base alla probabilità di abbandono, sia come stima del tempo di permanenza del cliente.

Previsione. La previsione viene distinta dalla classificazione e dalla stima per il fatto che i record sono classificati secondo valori stimati futuri. La differenza con l'attività di stima risiede nella fase di verifica del risultato, poiché per la previsione il controllo è concepibile teoricamente solo in un momento futuro. Strettamente legati all'attività di previsione sono i modelli di analisi delle serie temporali, che analizzano i dati riferiti a un certo arco di tempo per identificare le tendenze future. Per esempio si possono costruire modelli predittivi legati ai comportamenti dei clienti attuali, al fine di anticipare il comportamento dei consumatori futuri e la loro propensione a rispondere ad annunci, offerte, promozioni o ad effettuare determinati acquisti.

Segmentazione e clustering. L'attività di segmentazione è essenzialmente basata sulle tecniche di cluster analysis che consentono di individuare gruppi omogenei, o tipologie, che presentano regolarità al loro interno in grado di caratterizzarli e differenziarli dagli altri gruppi. Dato un insieme di unità statistiche (ad esempio, una certa tipologia di clientela) per le quali si conoscono un certo numero di variabili, l'analisi cluster si propone di ricercare

dei sottoinsiemi che siano massimamente omogenei al loro interno e il più possibile diversi uno dall'altro, rispetto alle variabili di input. Una delle più importanti applicazioni di clustering consiste nella segmentazione della clientela acquisita e potenziale.

Raggruppamento per affinità e associazione. Si cerca di individuare regole non evidenti per associare i dati, collegando più eventi o fatti elementari a una relazione di causalità. Dati un certo numero di proprietà e un insieme di record, l'associazione è un'operazione che evidenzia le affinità tra le proprietà selezionate. Le regole associative mettono in relazione una particolare conclusione, come l'acquisto di un determinato prodotto, a un insieme di condizioni, come l'acquisto di altri prodotti in un paniere articolato. L'individuazione degli oggetti da associare è realizzata in modo da prevedere, per esempio, quali prodotti vengono acquistati congiuntamente o per evidenziare la relazione tra i diversi consumatori e i prodotti acquistati.

Visualizzazione. Affinché l'utente del sistema possa avere una visione globale e approfondita delle regolarità presenti nei dati analizzati, le informazioni ottenute devono essere rappresentate. L'attività di visualizzazione nel data mining intende dare visibilità ai dati per una loro migliore interpretazione, può avere forme e rappresentazioni diverse ed è utile per individuare particolari distribuzioni dei dati. Gli strumenti di visualizzazione sfruttano le comuni doti di percezione come metodo di analisi basandosi sul concetto che, ciò che i numeri non possono dire, può spesso essere rivelato dall'immagine corrispondente. Lo scopo dei sistemi che utilizzano la tecnica di visualizzazione è ottenere facilmente e velocemente la rappresentazione visuale di grandi database da differenti prospettive e di scoprirne nuove e interessanti proprietà, correlazioni ed eventuali deviazioni dai valori attesi.

Ciò che distingue l'attività di data mining da un'analisi statistica tradizionale, quindi, non è tanto la quantità dei dati analizzati o la disponibilità di un numero rilevante di tecniche, quanto la necessità di operare integrando metodologie di analisi quantitativa e conoscenze di business. L'attività di analisi, perciò nasce ed è finalizzata alle esigenze aziendali, ed è svolta in un ambiente fortemente predisposto per l'integrazione tra contributi tecnici e di sistemi informativi.

2.2

Tecniche di indagine statistica

Le metodologie e le tecniche statistiche necessarie per una completa applicazione di data mining si dividono in *descrittive* e di *verifica*.

Le *descrittive* identificano relazioni e tendenze nei dati, aiutando a scoprire fenomeni di mercato e di conseguenza, a consolidare le conoscenze di base sul proprio business. Le nuove conoscenze possono essere usate per aumentare i propri margini di competitività o il proprio giro d'affari.

Nell'ambito delle tecniche descrittive si annoverano¹:

l'indagine esplorativa: vede la realizzazione di sintesi delle distribuzioni secondo indici di posizione e di variabilità oltre a rappresentazioni grafiche dei fenomeni;

gli alberi decisionali: sono rappresentazioni grafiche basate su ripetute suddivisioni dei dati secondo i risultati di alcune variabili in analisi con l'obiettivo di definire gruppi omogenei;

¹ Landenna G., *Fondamenti di statistica descrittiva*, Il Mulino, Bologna, 1994.

le reti neurali: modelli che grazie ad in processo di apprendimento, modificano i propri parametri interni fornendo soluzioni a problemi di classificazione e di previsione.

cluster analysis: permette di definire tipologie di unità statistiche massimamente omogenee al loro interno e massimamente eterogenee tra loro in funzione delle variabili di analisi;

analisi fattoriale: consente di ridurre lo spazio delle variabili restituendo fattori di sintesi a patto di cedere quota della variabilità spiegata del fenomeno.

Se gli strumenti di indagine possono portare a interessanti scoperte sulla attività di business, non possono, però, spiegare perché e se queste scoperte sono valide e utili. Le tecniche di verifica, invece, servono proprio a convalidare le scoperte fatte in sede di indagine, per garantire decisioni corrette. Alcuni degli strumenti di verifica sono:

le correlazioni: misura dell'associazione lineare tra due variabili. Il valore assoluto del coefficiente di correlazione indica l'intensità della relazione, mentre il segno del coefficiente indica la direzione della relazione;

il test T: la procedura viene applicata a un campione o a gruppi di campioni e consente di verificare se la media di una variabile è diversa da una costante specificata;

ANOVA: l'analisi della varianza è un modo per verificare l'ipotesi nulla che le medie di gruppo sono uguali nella popolazione, confrontando la varianza campionaria stimata dalle medie di gruppo con quella stimata entro i gruppi;

la regressione lineare: consente di stimare i coefficienti dell'equazione lineare, comprendenti una o più variabili indipendenti, che prevedono al meglio il valore della variabile dipendente;

l'analisi discriminante: utile in situazioni in cui si desidera creare un modello di previsione del gruppo di appartenenza in base alle caratteristiche delle

funzioni osservate; tali funzioni vengono generate da un campione di casi di cui è noto il gruppo di appartenenza;

le previsioni: rappresentano una stima, puntuale o per intervallo, dei futuri valori che il fenomeno potrà assumere, costruita sulla base di valori passati ed attuali sotto forma di modello.

In generale, attraverso un processo di data mining si ha la possibilità di comprendere meglio i dati di cui un'azienda dispone, mentre le tecniche di indagine statistica permettono di riassumere i dati storici per capire il passato e per prendere decisioni per il futuro, di segmentare il mercato per ottenere programmi più specifici, di assegnare un punteggio ai clienti esistenti per valutare i potenziali a breve termine ed ottenere un ritorno sugli investimenti.

2.3

Eventuali ambiti applicativi

Per rispondere alle nuove esigenze create dalle trasformazioni tecnologiche e di mercato, numerose sono, in ambito economico-aziendale, le applicazioni del data mining da quelle più tradizionali, relative alla funzione marketing, a quelle più recenti e innovative collegate al mondo di internet e dell'e-commerce.

In seguito si riporta una trattazione più dettagliata di alcune tra le più comuni applicazioni di data mining facendo specifico riferimento all'area del

marketing e delle ricerche di mercato in cui tale tecnica ed appropriate indagini statistiche sui dati raccolti, sono principalmente utilizzate al fine di ottenere informazioni utili al miglioramento delle proprie strategie competitive.

2.3.1

Valutazione degli effetti delle promozioni di vendita

Chi acquista un prodotto o un servizio è un soggetto chiamato a prendere numerose decisioni, il cui grado di incertezza varia soprattutto in ragione della natura del bene e della complessità della situazione d'acquisto. Il problema riguarda essenzialmente i meccanismi che determinano le scelte dei consumatori e può essere analizzato attraverso l'osservazione di tutte quelle azioni che precedono, accompagnano e seguono le decisioni di acquisto.

A tal proposito, di fondamentale influenza sui comportamenti di un consumatore sono le promozioni di vendita. Una strategia promozionale è formata da una serie di elementi che vengono combinati nell'arco del periodo di riferimento e che considerati complessivamente fanno in modo che essa abbia un determinato peso. È, dunque, profondamente diverso valutare una singola promozione rispetto a una serie successiva di promozioni applicate regolarmente e facenti parte di un preciso obiettivo di marketing.

Una strategia promozionale può essere formata da quattro differenti caratteristiche².

Tipo di offerta promozionale: si tratta di un indice sintetico che descrive il tipo di promozioni effettuate da ciascun punto vendita nell'arco del periodo considerato. Tra le tipologie promozionali maggiormente utilizzate è possibile annoverare la promozione *multibuy of identical item* (i cosiddetti 3 X 2 e simili), i *gift* (omaggio di un campione di prodotto), i *coupon* (buono sconto su ulteriori unità di prodotto) e la *promozione manufacturer/retailer price offer* (riduzioni di prezzo operate dal produttore/distributore).

Contemporaneità fra promozioni differenti: nel senso che possono essere impiegate più promozioni differenti nella stessa settimana. L'idea di base è che due o più promozioni attuate contemporaneamente permettono di ottenere risultati di vendita migliori.

Distanze tra periodi promozionali: stanno ad indicare il periodo di pausa, cioè le settimane in cui non vengono praticate promozioni. Il principio di fondo è che la definizione di tali intervalli permette di massimizzare l'efficacia del periodo promozionale. Le distanze tra periodi promozionali possono essere:

- *regolari:* le pause risultano simili e il consumatore è spinto all'acquisto nel periodo promozionale, mentre in quello di pausa tende ad acquistare solo il necessario sapendo che la promozione verrà ripetuta. Ciò permette al consumatore di restare fedele al prodotto o alla marca e di non essere invogliato a cercare altri generi simili in offerta;
- *irregolari:* gli intervalli tra un periodo promozionale e l'altro sono differenti. Il consumatore, allora, tenderà ad acquistare solo nel periodo

² Brasini S. et al., *Statistica aziendale e analisi di mercato*, Il Mulino, Bologna, 2002.

promozionale, mentre durante la pausa sarà portato ad abbandonare il prodotto per cercarne altri simili in offerta;

- *anomale*: le pause promozionali sono talmente brevi da non permettere distinzioni tra periodi promozionali differenti. Di conseguenza il consumatore vedrà la promozione come una situazione abituale e nel momento in cui questa sarà sospesa egli abbandonerà il prodotto.

Combinazioni tra caratteristiche: nel periodo considerato si analizza l'eventuale concomitanza fra distanze e contemporaneità nelle promozioni. Un mix tra caratteristiche positive genera un differenziale competitivo rispetto ad altre strategie prive di tali combinazioni.

Dal lato dell'azienda che mette in atto tali "offerte promozionali", invece, un obiettivo molto importante è la previsione sull'andamento futuro delle vendite perché fornisce sostanzialmente le informazioni di base per organizzare nel migliore dei modi le risorse disponibili e tutte le funzioni aziendali: dagli acquisti alla produzione, al personale, alla commercializzazione e gestione finanziaria. In questo contesto assume un'importanza cruciale la **previsione scientifica**, la cui specificità va individuata sia nel modo di osservare le cose esistenti sia nella deduzione su questa base di eventi futuri secondo regole generalmente riconosciute.

I metodi di previsione di carattere statistico sono basati generalmente su modelli che assumono come valide, per il futuro, relazioni che si sono verificate in passato e che, in una fase successiva, possono essere utilizzate per formulare delle ipotesi su alcuni dei possibili cambiamenti di condizioni alla base delle relazioni stesse. Da un punto di vista operativo, per prevedere ci si basa sulle informazioni disponibili, quindi sulle informazioni del passato e al più del presente; informazioni che se memorizzate su supporti informatici, possono essere estratte e rielaborate con le tecniche di data mining.

Uno dei metodi maggiormente utilizzati per le previsioni è il **metodo degli scenari** con il quale si vanno ad esaminare sistematicamente le possibili conseguenze di alcune delle situazioni future e consente di individuare le condizioni per il loro verificarsi. Tale metodo si distacca dalla semplice estrapolazione di tendenze passate e punta allo studio di un set di futuri plausibili determinati dal diverso possibile evolversi delle variabili chiave: non si risolve, quindi, in una sola previsione, ma in un set di previsioni possibili.

Fra i metodi statistici utilizzabili per lo studio delle tendenze, rivestono particolare importanza i **modelli lineari e non lineari di regressione semplice e multipla**³. Tali metodi partono dal presupposto che una volta individuati i fattori o le cause che si ritiene agiscano su una data variabile (ad esempio, sulle vendite di un dato prodotto) si possa puntare alla specificazione della relazione tramite l'individuazione di un'equazione o di un sistema di equazioni che legghi le variabili indipendenti a quella dipendente. L'obiettivo è di giungere, se non a vere e proprie generalizzazioni, almeno ad una buona approssimazione delle leggi statistiche che regolano il comportamento delle variabili in esame.

Prendendo, per esempio, in considerazione un prodotto maturo, con vendite e consumi tendenzialmente stabili, è tuttavia plausibile l'ipotesi che la domanda possa risultare moderatamente espandibile a seguito del lancio di un prodotto con carattere di novità e la conseguente pressione di campagne pubblicitarie, promozionali e politiche di prezzo tese a stimolare la tensione all'acquisto del nuovo prodotto.

La prima operazione da effettuare riguarda la precisazione dell'obiettivo da raggiungere che è quello di valutare l'influenza dei singoli fattori sulle vendite e di prevedere, sulla base delle loro possibili variazioni, la domanda futura. A questo proposito il principale problema da risolvere è quello di stabilire come

³ Piccolo D., *Statistica*, Il Mulino, Bologna, 2000.

possono essere individuati quei fattori ritenuti importanti per spiegare le variazioni della domanda; problema che può trovare soluzioni adeguate solo se si considerano non superficialmente il mercato a cui la domanda viene rivolta, e in particolare, sul versante dei potenziali acquirenti, le loro preferenze e gli inevitabili condizionamenti.

Il problema successivo è quello della misura delle variabili, la cui soluzione trova un sostegno importante nella conoscenza delle fonti statistiche disponibili. Ma una volta estratte le informazioni, resta aperto l'interrogativo riguardante la loro qualità.

Come già accennato, uno degli strumenti statistici disponibili per la previsione delle vendite è il **modello di regressione** che consente di analizzare la relazione di dipendenza di tale variabile da un insieme di fattori che la influenzano. Il legame può essere descritto da una equazione (**equazione di regressione multipla**) del tipo

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_m X_m + \varepsilon$$

dove i simboli utilizzati indicano rispettivamente:

- Y la variabile dipendente, nel caso di specie la domanda;
- β_j il coefficiente o parametro del modello relativo al fattore j-esimo che misura l'influenza di X_j su Y;
- X_j il fattore j-esimo (numero dei fattori considerati);
- ε il termine d'errore.

È necessario completare l'espressione con un termine di errore perché si presume che in esso confluiscono gli effetti delle variabili non incluse nella

relazione, gli errori di misura delle variabili in esame e la componente di aleatorietà intrinseca nel fenomeno considerato.

A questo punto l'analisi prevede che si proceda alla stima dei coefficienti β della relazione. A tal fine si può ricorrere al metodo di adattamento dei minimi quadrati che garantisce le migliori stime possibili a parità di ampiezza del campione.

Ma, prima di prendere in considerazione le stime dei coefficienti, è necessario valutare anche se il modello è adeguato a descrivere ed eventualmente a prevedere le vendite. La verifica si basa sulla coerenza delle proprietà degli errori con le ipotesi e se non ci sono fattori rilevanti esclusi, il modello potrà essere considerato soddisfacente. Ciò significa che quanto più i fattori sono rilevanti tanto più gli errori sono piccoli e il modello è adeguato a rappresentare il fenomeno considerato. Tuttavia, da un punto di vista statistico, è da notare che quanto più un modello descrive perfettamente il comportamento osservato è da considerarsi errato perché è un modello che descrive anche la parte del comportamento dei soggetti economici che dipende dal caso. Di conseguenza l'inutilità di questa descrizione sarà evidente in fase di previsione dato che la componente casuale non si manifesta quasi mai nello stesso modo. In altri termini, approssimare esattamente le osservazioni impiegando modelli molto complessi (con numerose variabili) porta a modellare anche comportamenti casuali che difficilmente potranno ripetersi tali e quali in futuro.

Infine, se le stime superano i controlli di significatività statistica, è possibile prendere in considerazione l'equazione di regressione per fare previsioni e calcolare i valori teorici di Y in corrispondenza dei valori di ciascuna X . Si tratta di disegnare scenari evolutivi per la domanda attraverso ipotesi alternative sulla determinazione futura delle variabili esplicative. Gli scenari possono essere determinati in maniera del tutto soggettiva sulla base dell'esperienza personale e della conoscenza delle dinamiche dell'economia.

In sostanza, con tali tecniche statistiche, è possibile constatare se le strategie promozionali influiscono in modo netto sulle vendite durante il periodo considerato. Per implementare correttamente queste strategie è però necessario tener conto delle caratteristiche e delle peculiarità di ogni classe di prodotto e adattare il modo di agire in riferimento alla specifica categoria.

È comunque necessario considerare anche eventuali altre variabili non promozionali che possono influire significativamente sull'andamento delle vendite. Esse possono essere per esempio i comportamenti di acquisto di un determinato Paese, l'ubicazione dell'attività, la posizione dei prodotti negli scaffali, le relazioni instaurate con i clienti. È evidente che i modelli previsionali potrebbero essere ulteriormente complicati inserendo tali nuove variabili esplicative, ma ciò potrebbe imporre l'adozione di metodologie previsive molto più complesse.

2.3.2

Metodi associativi per la *Market Basket Analysis*

Per l'analista di mercato rappresentano obiettivi cruciali sia individuare regolarità statistiche e profili di consumo da utilizzare come coordinate di fondo per successivi approfondimenti, sia per capire se eventuali mutamenti dei consumi di una popolazione sono espressione dell'emergere di bisogni nuovi a carattere più o meno duraturo. Per rispondere a queste esigenze è di fondamentale importanza la disponibilità di serie statistiche sufficientemente

lunghe, organizzate secondo schemi di classificazione delle singole voci di spesa.

Le aziende che operano sui diversi mercati sono per loro natura soggetti attivi, che si propongono in primo luogo di ottenere e consolidare nel tempo il gradimento e la preferenza dei consumatori per le proprie offerte. A tal fine mettono in atto principalmente strategie di marketing-mix che possono essere interpretate come stimoli sottoposti all'attenzione di tutti i consumatori potenziali; a seconda delle particolari reazioni fisiche o intellettive suscitate da tali stimoli, si possono poi identificare risposte differenti.

Nella risposta del consumatore vengono distinte abitualmente tre fasi che si susseguono in sequenza:

- **cognitiva:** che chiama in causa le conoscenze e le informazioni di cui dispone il singolo consumatore;
- **affettiva:** che si collega al sistema di preferenze specifiche del consumatore e alla sua valutazione soggettiva;
- **comportamentale:** che descrive le azioni, ossia gli atti di acquisto e il comportamento successivo.

Il processo di acquisto di un consumatore ha origine al momento dell'emergere di un bisogno, passa poi per le fasi di ricerca delle informazioni, di valutazione delle caratteristiche del prodotto, di decisione di acquisto e si conclude al momento della formazione delle sensazioni post-acquisto.

Il consumatore, in genere, svolge un attento esame delle alternative di prodotto qualora le differenze tra le marche risultino note ed evidenti, mentre prende in considerazione altri fattori, quali il prezzo, la localizzazione del punto vendita e la competenza del personale nel caso in cui non si rilevino differenze importanti.

Ai precedenti aspetti bisogna aggiungere l'influenza esercitata dalle politiche di marketing, che si riflette nel numero di marche esistenti, nel tipo di distribuzione, nelle campagne pubblicitarie effettuate, ecc. Le variazioni di prezzo e le attività promozionali e pubblicitarie sono, infatti, assimilabili a fattori esterni che condizionano temporaneamente i livelli delle vendite, senza però apportare effetti di lungo periodo.

Per un efficace sviluppo dei piani di marketing all'azienda giova conoscere il comportamento del consumatore anche nella fase successiva all'acquisto, nel breve come nel lungo periodo. È evidente, infatti, che il livello di soddisfazione conseguito, o eventuali stati di dissonanza suscitati dalla scelta effettuata, sono determinati in vista del comportamento futuro dell'acquirente o di altri potenziali acquirenti. Ciò giustifica l'interesse per la valutazione non solo delle vendite di primo acquisto, di un prodotto, ma soprattutto di quelle derivanti dal riacquisto.

Negli ultimi 20 anni in Italia si sono fortemente sviluppate organizzazioni commerciali complesse e di conseguenza si è verificato un processo di intensa concentrazione messo in atto dalla grande distribuzione. E proprio nell'ambito della grande distribuzione, la progressiva introduzione di tecnologie informatiche per la gestione dei punti vendita ha favorito la produzione di dati accurati e sistematici, che costituiscono il presupposto indispensabile per il ricorso ai metodi quantitativi nei processi decisionali relativi alle fasi del controllo direzionale ed operativo della gestione. Le aree interessate riguardano la programmazione dei livelli di servizio da offrire ai clienti, la gestione del magazzino, l'ampiezza e la profondità degli assortimenti, le politiche di prezzo, la valutazione dell'efficacia di iniziative promozionali messe in atto presso il punto vendita.

Come già accennato nelle grandi aziende di distribuzione le decisioni che coinvolgono le politiche di marketing possono essere in gran parte formalizzate e pianificate con il supporto di tecniche quantitative, a partire dal complesso di

dati resi disponibili grazie all'introduzione delle nuove tecnologie informatiche. L'esempio più importante, in questo ambito, è quello relativo ai cosiddetti sistemi POS (*Point of Sales*). Con tale espressione viene identificata la versione evoluta del registratore di cassa che, oltre all'emissione dello scontrino fiscale, permette di accedere ad un archivio interno (data warehouse) in cui vengono registrate per ogni prodotto, la marca, i prezzi pagati dai consumatori, le quantità vendute, nonché altre informazioni relative alle attività promozionali. L'efficacia di queste ultime viene misurata in termini di vendite aggiuntive attraverso un procedimento di stima del livello delle vendite di base, cioè quelle dovute esclusivamente alla presenza del prodotto sullo scaffale in assenza di qualunque attività di promozione. La differenza tra le vendite effettive e di base fornisce una stima delle vendite incrementalì a seguito di promozioni attive.

Sempre grazie alla disponibilità di dati rilevati continuamente sui punti vendita della grande distribuzione organizzata, si è andata affermando in questi ultimi anni una metodologia di analisi che va alla ricerca delle associazioni di acquisto più frequenti tra prodotti e marche.

Con il termine *Market Basket Analysis*⁴ vengono, infatti, considerati metodi statistici per l'individuazione di associazioni fra dati di vendita, al fine di conoscere quali prodotti sono acquistati congiuntamente ad altri; questo tipo di informazione consente di migliorare l'offerta dei prodotti e di incrementare le vendite di alcuni prodotti tramite offerte sui prodotti ad essi associati. L'idea di base di tale metodologia è che, per un consumatore, la performance di uno specifico prodotto possa avere un valore segnaletico che oltrepassa i confini della categoria a cui appartiene. Per comprendere meglio come opera la *Market Basket Analysis* si può prendere in considerazione il carrello della spesa di un cliente di un ipermercato. L'insieme dei prodotti di natura e impiego diversi contenuti in tale carrello, a ben vedere fornisce molte informazioni sulle

⁴ Giudici P., *Data mining: metodi statistici per le applicazioni aziendali*, op. cit.

abitudini di spesa del cliente. La *Market Basket Analysis* utilizza, dunque, come dati in input proprio quelli che provengono dagli scontrini fiscali a seguito di una scansione dettagliata dei prodotti acquistati, in modo da poter fornire al management le informazioni su chi sono i clienti e perché vengono effettuate certe tipologie di acquisti multipli.

Uno degli strumenti statistici che è possibile utilizzare per la l'analisi delle associazioni tra coppie di prodotti è il test χ^2 (**chi quadrato**)⁵ che consente di verificare l'ipotesi di indipendenza tra due variabili categoriche. Con tale test, quindi, si vuole vagliare la possibilità di incrementare le vendite mettendo in promozione solo determinati prodotti, al fine di vendere anche quelli ad essi associati.

Ai fini dell'applicazione di tale modello è necessario disporre delle frequenze di acquisto dei prodotti di cui si vuole analizzarne l'associazione.

L'ipotesi nulla e quella alternativa che vengono formulate sono:

- H_0 : le due variabili categoriche sono indipendenti;
- H_1 : le due variabili categoriche sono dipendenti.

La formula da applicare il χ^2 per calcolare è la seguente:

$$\chi^2_{(r-1)(c-1)} = \frac{\sum_i (n_{ij} - n_{ij}^*)^2}{n_{ij}^*}$$

dove

$r = n^\circ$ righe

$c = n^\circ$ colonne

per il calcolo dei gradi di libertà;

n_{ij} = frequenze osservate;

⁵ Piccolo D., *Statistica*, op. cit.

$$n_{ij}^* = \frac{n_{i0} * n_{0j}}{N} \text{ per il calcolo delle frequenze teoriche attese.}$$

Per accettare l'ipotesi nulla H_0 o quella alternativa H_1 è necessario confrontare il χ^2 calcolato con quello teorico:

- se il χ^2 calcolato $< \chi^2$ teorico l'ipotesi di indipendenza H_0 deve essere accettata;
- se il χ^2 calcolato $> \chi^2$ teorico l'ipotesi di indipendenza H_0 deve essere rifiutata.

Ottenere tutte le informazioni necessarie ad un'analisi market basket è comunque molto oneroso in termini di tempo e di risorse. Bisogna predisporre meccanismi che permettano di identificare i clienti e seguire l'evoluzione dei loro acquisti nel corso tempo, come ad esempio le *fidelity card* o la possibilità di trasferire i dati dalle casse direttamente all'interno di un data warehouse.

In ogni caso una volta appurata la presenza di una qualche relazione tra prodotti il primo passo è quello di riorganizzare il layout del punto vendita. Questa riorganizzazione deve puntare a raggruppare i prodotti in zone omogenee, nelle quali il consumatore non corra il rischio di dimenticarsi di comprare qualcuno dei prodotti che costituiscono il gruppo. Successivamente sarà necessario realizzare un piano razionale in cui si alternino le promozioni in maniera ciclica e in cui i due prodotti, di cui se ne è accertata l'associazione, non siano mai messi in promozione contemporaneamente. In questo modo l'acquirente è sollecitato all'acquisto dalla promozione di un prodotto, e la presenza delle associazioni con gli altri prodotti lo spinge ad ulteriori acquisti all'interno del gruppo.

2.3.3

Sondaggi di opinioni e ricerche di mercato per la *Customer Satisfaction*

Il costante aumento dell'attività competitiva conseguente alla globalizzazione dei mercati, la diffusione di nuove tecnologie e la flessibilità dei processi produttivi implicano per le aziende la necessità di individuare e soddisfare in modo più efficace rispetto ai concorrenti i bisogni tangibili e intangibili della clientela.

La nozione di *customer satisfaction*⁶ si fonda sostanzialmente su tre proposizioni:

- la qualità del servizio fornito dall'azienda deve tradursi in soddisfazione dei clienti;
- la soddisfazione aumenta la fedeltà della clientela, che si manifesta in acquisti ripetuti dello stesso prodotto;
- una clientela fedele assicura all'azienda profitti più elevati.

Per l'azienda, dunque, capire quale sia il grado di soddisfazione della clientela costituisce un momento conoscitivo di fondamentale importanza ai fini dell'acquisizione e del consolidamento nel lungo periodo di vantaggi competitivi.

L'esigenza di conoscere e valutare la qualità di un bene o di un servizio in termini di soddisfazione della clientela trova risposta nei risultati di ricerche

⁶ Lambin J. J., *Marketing*, Mc Graw-Hill, Milano, 1991.

sui meccanismi di natura economica e psicologica che ne stimolano l'acquisto. Per valutare le caratteristiche del prodotto così come sono percepite dalla clientela un approccio di notevole interesse è quello che prevede la raccolta di informazioni presso campioni di clienti sulla percezione della qualità che essi hanno del prodotto stesso; informazioni che opportunamente elaborate consentono la costruzione di indicatori del grado di soddisfazione raggiunto. Le operazioni e gli strumenti indispensabili per raggiungere questo scopo sono:

- la definizione di un *disegno campionario* per la selezione casuale dei soggetti da intervistare;
- la messa a punto di un *questionario* per la raccolta delle valutazioni della clientela;
- la scelta delle *modalità dell'intervista*;
- la scelta della *tecnica di analisi statistica* più idonea per l'elaborazione dei dati raccolti.

Il disegno campionario

Le informazioni statistiche di base per analizzare la soddisfazione, le preferenze e i comportamenti d'acquisto dei consumatori si ottengono attraverso l'osservazione di un sottoinsieme delle unità statistiche appartenenti alla popolazione oggetto di interesse.

Il ricorso alla **rilevazione campionaria** è giustificato dai risparmi di tempo e di mezzi conseguibili rispetto ad una rilevazione di tipo completo che risulterebbe molto più lunga, complessa e dispendiosa. Per assicurarsi informazioni di buona qualità, è però necessario che il campione osservato sia **rappresentativo**. Ciò significa che per quegli aspetti della popolazione ritenuti rilevanti ai fini dell'operazione di controllo si dovrebbe ottenere dal campione un'immagine abbastanza fedele della popolazione stessa.

La formazione di un campione è realizzabile seguendo approcci diversi, riconducibili alle fondamentali procedure della **scelta ragionata** e della **scelta casuale**.

Nel caso di *scelta ragionata*, la selezione delle unità da osservare è lasciata all'iniziativa dell'operatore e la dimensione del campione viene fissata di norma in base a criteri di pura convenienza economica e pratica.

Il *campione casuale*, invece, è soltanto uno fra tanti campioni possibili ed ha il grande vantaggio di eliminare quel soggettivismo pericoloso e difficilmente accettabile, tipico di un campione ragionato. Solo se il campione è scelto casualmente è possibile calcolare la probabilità dell'errore a cui si va incontro nella stima delle caratteristiche oggetto di interesse ed estendere i risultati all'intera popolazione mediante procedimenti di inferenza induttiva inversa: inversa perché dal particolare (il campione) si cerca di risalire al generale (la popolazione). Occorre aggiungere peraltro che se le unità che formano il campione sono selezionate senza alcun criterio particolare ma in numero sufficientemente ampio, tale da poter ritenere di cogliere per intero la variabilità del fenomeno che si sta studiando, la casualità del campione può essere presunta. L'inevitabile margine di errore che accompagna i risultati di un'indagine campionaria, richiede che la loro rappresentazione sia corredata da informazioni che ne consentano una lettura corretta.

La teoria degli intervalli di confidenza rende disponibile la strumentazione statistica necessaria per valutare il rischio di commettere errori casuali di prefissata ampiezza nella stima delle variabili di interesse. Il modo più semplice di formare un campione da una popolazione costituita da un numero finito (N) di elementi consiste nel selezionarne casualmente un numero prestabilito ($n < N$) in modo da garantirsi che tutti i possibili campioni abbiano la stessa probabilità di essere estratti. Tale procedimento, semplice ma efficace, è chiamato **campionamento casuale semplice o estrazione in blocco**.

Esistono comunque piani di campionamento alternativi che richiedono il ricorso a **procedimenti di stratificazione** e prevedono la possibilità di

suddividere la popolazione in gruppi o strati omogenei secondo criteri prestabiliti. La selezione del campione si realizza in due o più fasi mediante la scelta casuale di grappoli (cluster) di unità statistiche elementari e successivamente, nell'ambito di ciascun cluster, di un numero prefissato di unità elementari sempre con selezione casuale. Le ragioni che inducono ad utilizzare schemi più complessi sono la finalità di contenere la dimensione del campione (il che comporta minor lavoro e minori spese) e di ottenere migliori risultati come conseguenza dell'aumento della precisione delle stime rispetto ad un campione casuale semplice.

Il questionario

Il questionario è un insieme strutturato di domande volte ad acquisire informazioni. Prima di costruire il questionario è però necessario: definire con esattezza l'obiettivo conoscitivo dell'indagine, soprattutto in relazione alle ipotesi di lavoro; stabilire i criteri di selezione di un campione di rispondenti, dopo aver identificato la popolazione di riferimento per l'indagine; precisare le modalità di contatto con gli intervistati.

Le decisioni riguardanti il contenuto dei quesiti hanno a che vedere soprattutto con la natura delle informazioni da rilevare: a quale specifico avvenimento o circostanza si fa riferimento. Condizione indispensabile è l'impiego di una terminologia chiara e facilmente comprensibile da parte di tutti i rispondenti.

Per ridurre al minimo la possibilità di generare errori al momento della compilazione del questionario, è di fondamentale importanza l'ordine di presentazione dei quesiti. L'interpretazione e, di conseguenza, la risposta che l'individuo dà ad una domanda dipendono dal contesto creato dalle domande precedenti. Quelle di apertura devono essere il più possibile semplici, oggettive ed interessanti. Se gli intervistati non sono in grado di rispondere agevolmente, o non le trovano degne di attenzione, è verosimile che rifiutino di rispondere

alle successive. Le domande sulle caratteristiche economiche, demografiche e sociali degli intervistati o quelle che suscitano diffidenza e imbarazzo per motivi di riservatezza, vanno, invece, collocate nella sezione finale. L'inserimento di queste domande alla fine del questionario è motivato dall'ipotesi che il suo completamento, almeno parziale, e quindi la conoscenza dell'argomento e dello scopo della ricerca, possano ridurre le resistenze a fornire questo genere di indicazioni.

La modalità di intervista

Per quanto concerne, invece, la scelta delle modalità dell'intervista, esistono quattro tipologie principali di rilevazione: *le interviste personali, le interviste telefoniche, i sondaggi postali e i sondaggi via internet*. La scelta della modalità più idonea è strettamente legata al tema dell'indagine, alle caratteristiche dell'ambiente e della popolazione, ai vincoli di tempo e di spesa da rispettare.

Dal punto di vista pratico, a fronte di costi di reperimento delle informazioni sempre maggiori, nei sondaggi demoscopici e nelle ricerche di mercato su campioni numerosi trovano sempre più frequente applicazione i sistemi informatici. Durante le interviste personali e telefoniche le risposte vengono salvate direttamente in un database abbinato al questionario. Un software controlla passo dopo passo lo svolgimento dell'intervista e compie in tempo reale verifiche di completezza e di coerenza tra le risposte. Attraverso la definizione di un piano di elaborazione (data mining), quindi, il sistema può generare tavole di spoglio e di consultazione dei risultati in modalità random a seconda delle esigenze informative.

Anche internet sta entrando con forza nel mondo dei sondaggi e delle ricerche di mercato. A tutti gli effetti i sondaggi via internet possono essere visti come la naturale evoluzione tecnologica dei più tradizionali sondaggi postali. Nella

modalità attualmente più diffusa, infatti, le indagini prendono l'avvio con la spedizione di un messaggio di posta elettronica (e-mail), assimilabile ad una vera e propria lettera di presentazione della ricerca alle unità statistiche designate a far parte del campione. All'interno del messaggio viene data comunicazione dell'indirizzo web nel quale è presente, on-line, un questionario predisposto in formato elettronico. Le risposte vengono inserite direttamente dall'intervistato mediante un form che gli appare sul monitor e anche in questo caso sono automaticamente memorizzate in un database abbinato al questionario con tutti i vantaggi prima evidenziati.

La tecnica di analisi statistica

Ultimo punto, ma di fondamentale importanza, concerne l'individuazione della tecnica di analisi statistica più idonea per l'elaborazione dei dati raccolti.

Numerose sono state le ricerche per l'individuazione di uno schema univoco di analisi per la *customer satisfaction*, ma la tecnica che si è dimostrata più adatta è l'**analisi fattoriale**⁷. Lo scopo di tale analisi è quello di condensare l'informazione contenuta in un numero elevato di variabili originarie in un numero esiguo di nuove variabili (*variabili latenti*), ottenute come combinazione lineare delle variabili di partenza, con una perdita minima di informazione. Il grado di soddisfazione della clientela per un dato prodotto può essere immaginato appunto come una di queste variabili latenti. Infatti, come spesso accade, il fenomeno comportamentale oggetto di studio è caratterizzato da un numero elevato di variabili che, almeno in parte, duplicano il medesimo contenuto informativo.

Nel caso della *customer satisfaction* alcuni dei fattori latenti su cui è possibile restringere il campo di indagine possono ad esempio riguardare il valore

⁷ Corbetta P., *Metodologia e tecniche della ricerca sociale*, Il Mulino, Bologna, 1999.

percepito, la qualità percepita, le attese della clientela i reclami, la fedeltà alla marca, il numero di reclami.

Si supponga di aver osservato un insieme di p variabili quantitative presso n unità statistiche il cui elemento generico x_{hj} denota il valore della variabile X_j osservato presso l'unità h . Il modello matematico per l'analisi fattoriale si basa sull'idea che il contenuto informativo di ciascuna variabile possa essere scomposto in due parti:

1. una parte di variabilità comune (sintetizzata nei fattori comuni);
2. una parte variabile (sintetizzata nei fattori specifici).

Il modello che ne deriva si esprime con l'equazione:

$$X_j = a_{j1} F_1 + a_{j2} F_2 + \dots + a_{jq} F_q + u_j c_j \quad (j = 1, \dots, p)$$

dove

- F_i rappresenta il fattore comune i -esimo;
- a_{ij} è il coefficiente che lega il fattore F_i alla variabile X_j e consente, applicato al relativo fattore, di riprodurre la quota di variabilità comune contenuta in ciascuna variabile;
- c_j è il fattore specifica di X_j e u_j il suo coefficiente.

Punto di partenza per la costruzione del modello, dato un insieme di osservazioni relative alle singole unità di un campione, è la matrice di varianze e covarianze, che ne esprime la variabilità complessiva. Punto di arrivo sono, invece, i coefficienti di equazioni strutturali che descrivono le relazioni tra le variabili.

In definitiva, l'esigenza di tenere costantemente sotto osservazione la clientela attuale e potenziale è collegata alla opportunità di ottenere indicazioni adeguate

per mettere a punto strategie di intervento. L'indagine sul consumatore non si configura, quindi, come un puro esercizio di ricerca, ma porta il suo contributo ad un sistema della qualità che vede, come ulteriori passaggi, la definizione delle strategie aziendali, la realizzazione delle stesse ed il monitoraggio delle performance di servizio offerto.

2.3.4

La gestione della qualità: il *Total Quality Management*

Il grande sviluppo della produzione di beni e servizi di largo consumo, resi disponibili a tutti ed in ogni luogo, a caratteristiche e prezzi sempre più competitivi, la grande diffusione dell'informazione, delle conoscenze e delle tecnologie, la crescita del livello di istruzione, hanno determinato oggi nel consumatore e nel mercato in generale, un'opportunità di selezione, di crescita e di diversificazione delle esigenze. Chi offre un bene od un servizio deve, quindi, cercare di soddisfare le aspettative dei propri clienti e deve impegnarsi al massimo per conservare a lungo la loro soddisfazione sia per evitare che il cliente insoddisfatto possa rivolgersi alla concorrenza che per evitarne lamentele e richieste di indennizzo. Tali nuove esigenze impongono alti livelli di know-how tecnico ed organizzativo, ma soprattutto alti livelli dell'efficienza delle azioni svolte e di uso delle risorse per ridurre al minimo gli errori e, quindi, gli sprechi e le perdite non solo economiche ma anche quelle che si misurano come mancato sfruttamento delle opportunità.

Il controllo delle attività, dei prodotti e dei servizi erogati, dunque, non è più possibile solamente attraverso attività di misura e limitando il controllo alle attività produttive, ma deve essere ripensato nei suoi principi di fondo, nei metodi e nelle tecniche e, soprattutto, negli ambiti di applicazione. In tale contesto, la qualità deve assumere valore sia in termini di contenuti di prodotto/servizio che come capacità di un'organizzazione di essere in grado di realizzare quel prodotto o di erogare quel servizio al meglio, rispondendo cioè, in modo ottimale, a predefinite esigenze.

Pertanto, se tradizionalmente il fattore qualità si basava soprattutto sull'assenza di difetti, oggi è diventato uno degli aspetti cruciali dell'intera attività aziendale in un contesto che viene etichettato come Qualità totale (*TQM - Total Quality Management*)⁸. In particolare con l'uso del termine "Totale" si vuole intendere, appunto, l'estensione della qualità alle più importanti funzioni aziendali fino a considerare il TQM un vero e proprio modo di governare l'azienda.

Gli ambiti applicativi più frequenti del controllo della qualità sono quelli riguardanti la verifica delle caratteristiche dei prodotti secondo un approccio che consiste nel sorvegliare i fattori di variabilità dei processi di fabbricazione che influiscono sulla qualità dei prodotti finiti, per assicurarsi che la produzione risponda a norme prefissate. In pratica, si tratta di mettere a punto una serie di definizioni e di misure a cui viene dato il nome di *specifiche*. Si dice, allora, che la qualità del prodotto consiste nel suo grado di conformità a tali standard, misurata in base alla percentuale di pezzi difettosi riscontrata al momento dell'uscita del prodotto dal processo di lavorazione. Nel controllo di lavorazione, dunque, in presenza di irregolarità produttive vi è l'esigenza di ottenere informazioni rapide e gli strumenti utilizzati per monitorare le singole fasi del processo produttivo sono le cosiddette carte di controllo. Tali strumenti servono a mettere in evidenza le differenze tra i valori rilevati e quello ideale

⁸ Tamburrano E., *Il controllo Statistico di qualità*, Franco Angeli, Milano, 1991.

fornendo su questa base le necessarie indicazioni per interventi di tipo correttivo necessari a riportare i valori fuori quota nei limiti tollerati. L'obiettivo di fondo è quello di distinguere tra le differenze imputabili al caso e quelle dovute a variazioni significative nel modo di produrre.

Ma, come già anticipato, il campo di applicazione della misura della qualità riguarda sempre più non solo il settore dei beni, ma anche quello dei servizi. Tuttavia, le caratteristiche di intangibilità, eterogeneità, e inseparabilità dell'erogazione che connotano i servizi e le difficoltà di individuare gli elementi di valutazione da parte del cliente, rendono la misurazione della qualità dei servizi un'operazione complessa.

Il tema della qualità dei servizi, emerso più tardi rispetto al contesto industriale, si è prevalentemente connotato sulla base di osservazioni empiriche, assumendo sempre come punto di vista il consumatore. In questo senso è stata approfondita soprattutto l'analisi di quegli elementi che determinano una buona percezione della qualità del servizio e, in particolare, la relazione tra aspettative e performance percepite dai clienti. In altri termini, la percezione della qualità del servizio da parte dei consumatori risulta dalla comparazione delle loro attese, cioè tra quello che essi considerano dover essere il servizio offerto e la loro percezione della performance effettiva.

Dunque, le aspettative rappresentano lo standard di riferimento da cui ha origine il processo valutativo del cliente. Dal confronto tra performance percepita e standard atteso si determina una discrepanza positiva/negativa dalla quale dipende a sua volta il giudizio di soddisfazione/insoddisfazione del cliente.

Pertanto, un'organizzazione dovrebbe acquisire competenze adeguate non solo per la costruzione di strumenti di controllo necessari alla verifica della qualità, ma anche per la predisposizione di strumenti informatici idonei per

l'archiviazione dei dati rilevati. Da questo punto di vista gli aspetti di maggior rilievo sono:

- la selezione dalle specifiche a cui applicare il controllo;
- la scelta di un appropriato metodo di controllo;
- la fissazione della frequenza di campionamento;
- il calcolo dei limiti di accettabilità (tolleranze);
- la raccolta dei dati su strumenti informatici di archiviazione (data warehouse);
- il calcolo degli indicatori di rilevanza statistica;
- l'esame dei risultati ottenuti e implementazione delle necessarie azioni correttive.

Tra le modalità statistiche che potrebbero essere utilizzate da un'organizzazione per tenere sotto controllo il proprio sistema di qualità, sono di primaria importanza il calcolo della varianza e la sua analisi (ANOVA)⁹.

Nella pratica sperimentale l'ANOVA (Analysis of Variance) ha sempre trovato ampia applicazione in quanto consente di ottenere una valutazione quantitativa dell'importanza delle diverse fonti di variazione nella variabilità osservata nel corso di un esperimento. In altri termini, l'ANOVA consente di scomporre la varianza di una variabile casuale in componenti indipendenti e relative a due o più livelli di uno o più fattori di variabilità e all'errore sperimentale.

Detta Y la variabile casuale mediante la quale vengono espressi i risultati sperimentali, un generico dato può essere rappresentato mediante la seguente equazione:

$$y_{ij} = \mu + \alpha_i + \varepsilon_{ij}$$

⁹ Piccolo D., *Statistica*, op. cit.

in cui:

- y_{ij} è la generica osservazione della variabile casuale Y ;
- μ è la media generale della popolazione da cui è stata estratta l'osservazione y_{ij} ;
- $\alpha_i = \mu_i - \mu$ è lo scostamento dalla media della i -esima popolazione dalla media generale μ ;
- ε_{ij} è il residuo o errore sperimentale, cioè la conseguenza di tutti i fattori che possono influenzare l'unità sperimentale e non sono o non possono essere valutati o controllati con l'esperimento.

Tale modello statistico è chiamato ANOVA ad un solo criterio di classificazione, ma in pratica può essere utile prendere in considerazione più di una fonte di variabilità (ad esempio β_j) dato che in molti casi è necessario verificare l'importanza relativa di due o più cause che agiscono sulle unità sperimentali, oltre all'eventuale interazione tra i fattori considerati $(\alpha\beta)_{ij}$.

Tuttavia, il controllo statistico della qualità è solo uno dei possibili campi di applicazione dell'analisi della varianza che invece trova applicazione in moltissimi contesti sperimentali e permette di ottenere, a seconda della finalità dell'esperimento:

1. la stima simultanea della precisione delle medie dei trattamenti o delle differenze tra di esse;
2. la verifica dell'ipotesi di uguaglianza tra le medie dei trattamenti e dell'assenza di interazione tra i fattori sperimentali quando essi sono più di uno;
3. la stima dei componenti della varianza totale attribuibili a diversi fattori presi in considerazione i cui livelli sono considerati come campionati da una popolazione o non direttamente controllati dallo sperimentatore.

2.3.5

Misura dell'audience e dell'efficacia della pubblicità

La pubblicità costituisce la parte più appariscente delle attività di marketing, anche se in realtà è solo una componente dell'offerta aziendale insieme agli attributi del prezzo, al prodotto, alle caratteristiche della distribuzione e alla promozione.

La pubblicità, così come definita da Lambin¹⁰, è *“L'azione che stimola il comportamento d'acquisto suscitando quell'atteggiamento favorevole che aumenta la probabilità di acquisto della marca (o del prodotto)”*.

Per spiegare il ruolo informativo della pubblicità, si può partire dal presupposto che il consumatore non ha informazioni complete sui prodotti, sulle marche e talvolta sulla loro stessa esistenza. Per essere efficace, tuttavia, un'azione pubblicitaria va inserita all'interno di una strategia globale di comunicazione con compiti ed obiettivi ben definiti, quali:

- suscitare il bisogno di un dato prodotto;
- diffondere la conoscenza e la consapevolezza degli attributi della marca;
- migliorare il giudizio o la preferenza dei consumatori rispetto alla marca;
- stimolare l'intenzione d'acquisto;
- portare a conoscenza dei potenziali acquirenti l'esistenza di condizioni che rendono l'acquisto particolarmente vantaggioso.

¹⁰ Lambin J. J., *Marketing*, op. cit.

Va sottolineato che quando si fa riferimento al comportamento di acquisto è spesso difficile separare l'effetto della pubblicità da quello degli altri fattori che influenzano l'acquisto stesso: la qualità percepita dall'acquirente, la sua precedente esperienza di consumo, la disponibilità nel punto vendita delle marche concorrenti, le campagne promozionali ecc. D'altra parte le vendite possono non manifestare apprezzabili variazioni in concomitanza o a seguito di ottime campagne pubblicitarie, proprio per l'effetto preponderante degli altri fattori.

Per superare le difficoltà che si presentano quando si cerca di separare l'effetto della pubblicità dagli altri fattori che contribuiscono a determinare il comportamento di acquisto si può fare ricorso a due strategie:

1. la realizzazione di esperimento di mercato;
2. la specificazione e la stima di modelli di risposta.

Negli esperimenti di mercato differenti livelli degli strumenti di marketing sono impiegati su diverse unità di studio (aree geografiche, negozi, famiglie, ecc.) per misurarne l'effetto specifico. Questo approccio sconta il limite di dover sottoporre a verifica empirica un numero limitato di livelli delle variabili di marketing, altrimenti il tempo necessario per concludere l'esperimento diventa eccessivamente lungo.

I modelli di risposta, invece, sintetizzano il modo in cui la variabile che esprime la performance dell'azienda (o del prodotto o della marca) reagisce ad altre variabili che rappresentano gli strumenti di marketing. Come variabile che rappresenta la performance si può fare riferimento alle vendite o alle quote di mercato.

Lo studio della risposta comportamentale non può prescindere dal considerare che gli effetti dell'azione pubblicitaria, da un lato, non si raggiungono immediatamente, ma solo dopo che un certo numero di messaggi è stato

percepito e, dall'altro, che solitamente si protraggono nel tempo; la risposta comportamentale è quindi un fenomeno intrinsecamente dinamico, i cui aspetti più rilevanti sono quelli degli effetti cumulativi.

Il problema di individuare il modo più efficace per fornire il numero desiderato di esposizioni del messaggio al pubblico, che costituisce l'obiettivo di un investimento pubblicitario, trova soluzione in un'adeguata selezione dei canali pubblicitari.

In questo contesto, con l'espressione numero desiderato di esposizioni si indica la combinazione di tre distinti fattori¹¹:

1. il numero di persone (o di famiglie) esposte al messaggio almeno una volta durante un dato periodo di tempo (**contatti**);
2. il numero di volte, sempre nello stesso intervallo temporale, in cui in media la persona è esposta al messaggio (**frequenza**);
3. il valore qualitativo dell'esposizione al messaggio attraverso un dato canale pubblicitario (**impatto**).

La pubblicità è applicabile ad un'ampia gamma di situazioni e per ciascuna di queste si può considerare una varietà di strategie e di mezzi di comunicazione di massa utilizzabili in relazione agli obiettivi (soprattutto di vendita) che l'azienda intende perseguire.

La scelta dei canali pubblicitari richiede la conoscenza:

- delle abitudini di lettura, di ascolto radiotelevisivo, di navigazione internet delle persone a cui si rivolge la campagna pubblicitaria;
- dell'impatto dei singoli mezzi di comunicazione;
- del costo unitario di ogni passaggio pubblicitario.

¹¹ Brasini S. et al., *Marketing e pubblicità*, Il Mulino, Bologna, 1999.

Su queste basi, prefissato l'ammontare dell'investimento che l'azienda è disposta a sostenere, è possibile individuare il mezzo pubblicitario più adatto e la combinazione più efficiente in termini di obiettivi di comunicazione tra numero di esposizioni e frequenza che minimizza il costo medio unitario del numero desiderato di esposizioni.

Audience di lettura

Alle ricerche sulla lettura si chiede di stimare non solo quanti lettori una testata può raggiungere, ma anche di fornire informazioni di carattere socio-demografico e comportamentale sui lettori stessi.

Dal punto di vista dell'inserzionista, le ricerche sui mezzi di comunicazione consentono di stimare la probabilità di esposizione del lettore ad una testata e traggono origine dalla necessità di informazioni per pianificare le campagne pubblicitarie. Gli editori, invece, sono interessati a queste ricerche per mostrare ai potenziali inserzionisti i vantaggi competitivi che derivano dal ricorso alle proprie testate.

Le indagini sono basate su un campione casuale di lettori ed hanno come obiettivo la definizione dell'audience di testata, vale a dire, il numero di lettori di un generico numero della pubblicazione. L'audience, così definita, corrisponde alle persone che hanno l'opportunità di leggere la pubblicità contenuta nella copia di testata.

Per l'individuazione di un lettore vengono utilizzate principalmente due tecniche: una basata sul ricordo da parte dell'intervistato di una precedente occasione di lettura della pubblicazione nel periodo di riferimento; l'altra basata sul riconoscimento, da parte dell'intervistato, di uno specifico numero letto nel periodo di riferimento.

I pericoli a cui vanno incontro le indagini sulla lettura sono determinati soprattutto da errori di memoria da parte dell'intervistato. Per attenuare questo

problema, il ricordo dell'intervistato può essere sollecitato da stimoli quali la copia della pubblicazione in oggetto, un cartellino riproducente la testata ecc.

Audience radiotelevisiva

Si definisce audience di un programma radiotelevisivo, il numero di persone che ne ha seguito una frazione almeno pari alla metà dell'intervallo temporale di riferimento. Si tratta anche in questo caso, come per le ricerche sulla lettura, di una definizione convenzionale, in quanto è necessario stabilire quale livello di attenzione deve essere prestato ad un programma affinché un individuo sia considerato spettatore/ascoltatore e possa essere incluso nell'audience.

I metodi di rilevazione dell'audience, nel tentativo di rispondere meglio ai cambiamenti intervenuti nel sistema radiotelevisivo, hanno subito nell'ultimo ventennio sensibili modificazioni. Gli strumenti elettronici di misurazione, a cui si ricorre attualmente, rilevano in modo oggettivo la sintonizzazione dell'apparecchio radio/televisivo ma non l'ascolto. Considerano, quindi, come spettatore chi è presente nella stanza in cui si trova l'apparecchio trasmittente.

Per valutare l'audience potenzialmente esposta ad un singolo spot, si può fare riferimento al numero di contatti ottenuti con un piano pubblicitario dato dalla somma delle audience riguardanti i programmi in cui le singole inserzioni pubblicitarie sono inserite. Un'altra misura di riferimento è data dal numero di persone raggiunte almeno una volta mediante il piano pubblicitario.

Per quanto riguarda i metodi di rilevazione bisogna distinguere tra quelli che fanno riferimento all'intervista diretta e quelli basati sull'accertamento automatico della sintonizzazione dell'apparecchio su di un dato programma tramite appositi registratori (*meters*). Mediante le interviste telefoniche si raccolgono informazioni relative al programma seguito ed ai singoli componenti della famiglia sintonizzati al momento della telefonata. I meters, invece, permettono la rilevazione automatica e continua dell'ascolto su un

panel di famiglie evitando di dar luogo ad un eccessivo carico per i rispondenti. In entrambi i casi le informazioni sono inviate ad un elaboratore centrale che le raccoglie in un data base e le rende disponibili per le necessarie elaborazioni.

Audience dei siti internet

Lo sviluppo di nuove tecnologie, in particolare la diffusione crescente di internet, sta producendo forti cambiamenti nello scenario dei mezzi di diffusione della pubblicità. Tramite internet, infatti, l'impresa è in grado di instaurare e gestire un rapporto di lungo periodo in un'ottica di rafforzamento della fedeltà alla marca. Gli operatori di marketing, considerando tali cambiamenti, dovranno probabilmente dotarsi nuove competenze, soprattutto per quanto riguarda la capacità di gestire e sviluppare relazioni con la clientela e di arricchire il contenuto della propria offerta interattiva.

Per la misurazione dei contatti con un sito web sono stati delineati due approcci:

- uno basato sull'analisi delle visite al sito attraverso tecniche di osservazione passive, ma il principale problema è che non si è in grado di conoscere le caratteristiche socio-demografiche del visitatore;
- un secondo approccio, economicamente meno conveniente, è basato sull'osservazione del comportamento del visitatore rilevato direttamente sul personal computer degli utenti che costituiscono il campione casuale d'indagine.

I dati così raccolti dai panels di utenti web, permettono di stabilire quali sono i siti internet più "cliccati", di analizzare il comportamento di visione e, conseguentemente, di individuare le strategie più efficaci di posizionamento delle inserzioni pubblicitarie on line.

Come per tutte le altre attività aziendali, anche per la comunicazione pubblicitaria è necessario mettere a punto strumenti statistici che possano essere utilizzati per misurare la congruenza tra obiettivi e risultati. A tal proposito, la vasta gamma degli obiettivi della comunicazione pubblicitaria fa sì che sussista un'ampia diversificazione sia dei dati necessari per esprimere una valutazione della sua efficacia, sia delle tecniche statistiche utilizzate per giungere ad una spiegazione razionale dei processi di risposta del consumatore.

Un modello di risposta descrive il modo in cui una variabile che esprime la performance dell'azienda reagisce ad altre variabili che rappresentano gli strumenti di marketing. Il principale e più immediato fra i modelli statistici è quello che si riferisce alla nozione di **reattività**¹². Tale concetto può essere espresso attraverso una funzione di **elasticità** in grado di quantificare la variazione della quota di mercato (o delle quantità vendute) conseguente alla variazione di una delle variabili del marketing mix.

Semplificando al massimo e supponendo che per influenzare la quota di mercato l'azienda possa utilizzare solo la pubblicità, l'elasticità può essere definita come segue:

$$e_{s_i, x_j} = \frac{\partial s_i}{s_i} \bigg/ \frac{\partial X_j}{X_j}$$

dove

e_{s_i, x_j} = elasticità della quota di mercato della marca i alla variabile X (pubblicità) della marca j ;

s_i = quota di mercato della marca i ;

$i = 1, \dots, m$ marche componenti il mercato di riferimento;

X_j = valore degli investimenti pubblicitari per la marca j .

¹² Mariani P., *La Statistica in azienda*, Franco Angeli, Milano, 1991.

La diffusione crescente dell'utilizzo in azienda di metodi statistici per lo studio della domanda trova ragione nella complessità dei mercati che rende impraticabile una conduzione del marketing basata sull'intuizione individuale e soggettiva. Un'impresa, infatti, per mantenere la sua posizione competitiva deve disporre di strumenti di comprensione del mercato che consentano di agire con rapidità, coerenza ed efficacia alle mosse dei concorrenti e della distribuzione. Anche il marketing e le ricerche dovranno pertanto assicurare a tutta l'azienda la disponibilità di strumenti cognitivi e previsivi certificati e continuamente aggiornati.

Come per qualsiasi altro investimento, ovviamente, la misura dell'efficacia è un'esigenza necessaria, anche se il contributo della sola variabile "pubblicità" non può essere stimato direttamente in quanto si tratta di uno tra i numerosi fattori che influenzano le decisioni di acquisto. Per valutare l'effetto della pubblicità sulle quote di mercato occorre, dunque, considerare l'insieme delle variabili di marketing influenti e stimarne il peso possibilmente in una prospettiva dinamica.

2.3.6

Analisi delle sequenze di visita ad un sito di *e-commerce*

Mentre le aziende utilizzano l'informatica per migliorare i propri sistemi informativi, l'innovazione tecnologica, informatica e telecomunicativa si è diffusa anche nella società: le famiglie si sono dotate di personal computer, l'uso di internet si è affermato e molti prodotti dall'editoria ai film hanno

assunto la forma digitale, creando opportunità inaspettate in ciascuno dei settori in cui tali prodotti sono stati sviluppati.

Anche il settore finanziario ha risentito fortemente di tale trasformazione. Grazie all'home banking si opera ormai da casa piuttosto che recarsi presso l'agenzia bancaria. Ma è forse nel commercio elettronico che ci sono stati i cambiamenti più vistosi sia dei consumatori che delle aziende. Via internet, gli ordini di acquisto vengono direttamente trasmessi ai produttori posizionati a monte della catena di produzione con notevoli risparmi in termini economici.

Il ritmo incalzante con cui l'innovazione tecnologica si sviluppa nel campo informatico e telecomunicativo rende problematico individuare le modalità di comportamento delle aziende e dei mercati in un futuro a medio termine.

Più precisamente la new economy si sta sviluppando con grande rapidità secondo tre direttrici:

- la crescita del numero di provider;
- la crescita del numero dei portali, con l'entrata sul mercato di grandi gruppi editoriali e industriali e la conseguente trasformazione di molti siti da "istituzionali" a fornitori di servizi sempre più articolati;
- il decollo dell'e-commerce.

È chiaro, quindi, che nuove modalità di indagine basate sul web godranno nei prossimi anni di un interesse crescente e non solo da parte di aziende operanti in nuovi mercati. Tali modalità, infatti, si inseriscono in maniera dirompente nella tradizionale dicotomia tra indagini personali e telefoniche, fornendo possibilità finora riservate alla prima tipologia di contatto con costi di ricerca delle informazioni addirittura inferiori alla seconda.

Un aspetto fortemente qualificante della ricerca sulla rete è costituito dai contenuti e dalle modalità di somministrazione delle interviste. Il web è un

media interattivo, che utilizza suoni immagini e filmati. Una ricerca su internet, perciò, può essere in grado di testare tutte le componenti di comunicazione che la rete offre, dall'immagine al filmato, dal sonoro alla simulazione di percorsi di navigazione.

Anche se oggi esistono molti software per l'effettuazione di ricerche sul web, alcuni dei quali in grado di gestire in modo efficiente tutti i supporti media, l'attività di ricerca sul web non è però, scevra da problemi tecnologici. Occorre, per esempio, prevedere una disponibilità di connettività sufficiente per realizzare la rilevazione; il server internet deve essere dimensionato in modo adeguato al fine di evitare che la connessione sia troppo lenta, o peggio che il numero eccessivo di collegamenti possa provocare la caduta del server stesso e la conseguente sospensione della rilevazione.

Attualmente, comunque, i problemi tecnologici non possono essere più considerati come un ostacolo rilevante al diffondersi delle indagini via web, perché, in ogni caso, tali problemi sono destinati a ridimensionarsi nel breve periodo.

Per quanto riguarda il web mining, ovvero l'analisi induttiva delle grandi quantità di informazioni generate dai contatti di un sito web, si tratta di un'area di interesse relativamente nuova e nella quale si stanno muovendo i primi passi.

Occorre precisare che per web mining¹³ non si intende la generazione di report, talvolta interessanti ma "banali" dal punto di vista statistico, quali il calcolo cumulato del numero di visite a ciascuna pagina del sito web di un'azienda. Con tale locuzione, invece, si fa riferimento alla ricerca di strutture nascoste dei dati che permettano di individuare percorsi di visita, relazioni tra caratteristiche del navigatore e comportamenti di navigazione e consentono,

¹³ Molteni L., *Nuove frontiere per le ricerche di marketing: web research e web mining*, in "Economia & Management", 1, 2001, pp. 42-43, SDA Bocconi, Milano.

conseguentemente, di prevederne pratiche di acquisto (se prevista la possibilità di e-commerce nel sito).

Più specificatamente, alcune domande chiave a cui il web mining può fornire risposta sono¹⁴:

- chi comprerà, cosa comprerà e quanto comprerà;
- quali sono le differenti tipologie di visitatori del sito;
- quali sono le relazioni tra prodotti/servizi proposti ai visitatori;
- come massimizzare la presenza on line e le vendite.

Le risposte a tali domande forniscono delle informazioni molto utili per un'analisi statistica volta all'individuazione della struttura associativa tra gruppi di pagine di un sito web che un navigatore visita e i suoi comportamenti di acquisto on line. Il problema è, pertanto, quello di esprimere la relazione tra questi due caratteri con un indice quantitativo e di valutarne la significatività e il grado di precisione.

A tal fine il metodo statistico più adatto alla misura del grado e del tipo di associazione è quello della **correlazione**.

Nel caso in esame, per misurare il grado di interdipendenza tra le due variabili considerate è possibile usare il **coefficiente di correlazione semplice lineare** di Bravais-Pearson¹⁵.

$$r_{xy} = \frac{S_{xy}}{S_x * S_y}$$

¹⁴ Giudici P., Castelo C., *Association models for web mining*, in "Journal of data mining and knowledge discovery", 5, 2001, pp. 183-196, Springer Netherlands.

¹⁵ Piccolo D., *Statistica*, op. cit.

dove

- S_{xy} è la **covarianza**, cioè la media aritmetica del prodotto dagli scarti di x e y che misura la contemporanea variazione di x e y in rapporto alle rispettive medie così calcolata:

$$S_{xy} = \frac{\sum_i (x_i - \bar{x}) (y_i - \bar{y})}{N}$$

- S_x e S_y sono le **deviazioni standard** di x e di y calcolate come radice quadrata delle rispettive varianze, quindi:

$$S_x * S_y = \sqrt{\text{Var}(x) \text{Var}(y)}$$

Per definizione $-1 \leq r \leq +1$:

- se $r = 0 \rightarrow$ non vi è alcuna relazione di tipo lineare tra i 2 caratteri analizzati;
- se $r = \pm 1 \rightarrow$ esiste un legame lineare perfetto di tipo concorde ($r = +1$) o discorde ($r = -1$);
- se $0 < r < +1 \rightarrow$ esiste una correlazione positiva incompleta;
- se $-1 < r < 0 \rightarrow$ la correlazione è negativa incompleta.

L'obiettivo di tale analisi statistica è quello di comprendere quali siano i comportamenti di navigazione e se, di conseguenza, vi sia una qualche possibilità che l'utente internet, visitando un sito, faccia i propri acquisti online. Ciò consentirà all'azienda di sfruttare al massimo le potenzialità di un sito di e-commerce, migliorando la struttura del sito stesso, indirizzando i visitatori verso le pagine desiderate e persuadendoli ad effettuare il maggior numero di acquisti possibili.

In sintesi, è possibile asserire che l'investimento in termini di tempi e di risorse sui temi del web mining e della web research, anche se rilevante, può essere

del tutto giustificato dalle sue potenziali ricadute sull'efficacia delle decisioni di marketing.

2.4

L'obiettivo: il CRM

La centralità del cliente per l'impresa ha assunto importanza crescente negli ultimi anni, al punto tale da far crescere intorno ad essa una vera e propria filosofia di gestione aziendale: il *Customer Relationship Management* (CRM).

Un comitato di esperti¹⁶ ha definito il CRM come:

“Una strategia aziendale per selezionare e gestire la clientela in modo da ottimizzare il valore di lungo termine. Il CRM richiede una filosofia e una cultura aziendale centrata sul cliente che supporti efficacemente i processi di marketing, vendite e assistenza. Le applicazioni CRM possono consentire la gestione di rapporti efficaci con il cliente, a condizione che l'azienda disponga della leadership, delle strategie e della cultura necessarie”.

Al centro del CRM c'è una predisposizione mentale, un insieme di processi e di politiche aziendali che coinvolgono tutta l'azienda, concepiti per acquisire, mantenere e servire la clientela. In generale il CRM include i processi rivolti al cliente relativi al marketing, alle vendite e all'assistenza.

Stabilire e mantenere relazioni, di lungo termine e reciprocamente utili con la clientela, deve rappresentare la ragion d'essere di un'organizzazione. Il cliente deve diventare un valore fondamentale dell'azienda.

¹⁶ Fonte: www.CRMGuru.com.

Il CRM deve cominciare con una strategia aziendale che determina cambiamenti nei processi organizzativi ed operativi che vengono resi possibili dalle tecnologie informatiche. Alla base dell'instaurazione e della gestione delle relazioni con i clienti, deve esserci per prima cosa uno sforzo strategico.

Il CRM come strategia aziendale non rappresenta una novità. Le aziende più attente hanno sempre capito di doversi concentrare sui clienti a più alto potenziale di vendite e profitti fornendo loro un buon servizio, per far sì che essi tornino ripetutamente.

La strategia per la creazione di una relazione con il cliente si applica nelle seguenti attività:

- *osservare che cosa il cliente sta facendo*; lo scopo è quello di accentrare tutta la conoscenza esistente sul cliente facendo confluire tutti i dati e le informazioni presenti nelle diverse aree aziendali e passanti dai vari canali. *“Il sistema di CRM deve essere gli occhi e gli orecchi dell'azienda e forse anche il naso, la lingua e le dita”*¹⁷;
- *ricordare che cosa i clienti hanno fatto nel passato*; essere cioè in grado di memorizzare le informazioni che si accumulano nel tempo e di organizzarle secondo una logica che mantenga il punto di vista sul cliente;
- *imparare da ciò che ci ricordiamo*; elaborare tali dati per trasformarli in informazioni comprensibili e utili per capire il cliente;
- *comportarsi conseguentemente a ciò che si è imparato per rendere i clienti più profittevoli*; una volta estrapolate le informazioni più rilevanti prendere delle decisioni in base a queste e agire con lo scopo di rendere i clienti più profittevoli.

¹⁷ Berry M., Linoff G., *Data mining techniques: for marketing, sales and customer relationship management*, Wiley, New York, 1997.

Ciò sta ad indicare che un cliente, quando si avvicina ad un'azienda, porta con sé un'aspettativa. Può trattarsi della necessità di un servizio o dell'interesse verso un prodotto ma, in ogni caso, esiste un'aspettativa che accompagna il suo interesse verso l'azienda.

Sarà l'esperienza vissuta a determinare il suo comportamento futuro. Una buona esperienza può aumentare la fedeltà e la tendenza al riacquisto fino a produrre la lealtà verso il marchio, mentre una cattiva esperienza può far allontanare il cliente dall'azienda ed innescare notevoli effetti negativi. La capacità di riconoscere questo processo e di gestirlo attivamente rappresenta la base del CRM.

Ognuno di noi sperimenta il CRM tutti i giorni. Andare dal parrucchiere preferito o portare l'auto a riparare dal meccanico di fiducia costituiscono interazioni con aziende che rappresentano "un'esperienza". La nostra propensione a rivolgerci di nuovo a quelle aziende sarà condizionata dal fatto di aver conservato ricordi positivi o negativi da queste esperienze.

Il punto non è se il CRM è importante per l'azienda, ma come utilizzarlo al meglio. La capacità di garantire all'interno dell'azienda un'azione unitaria continuativa ed un comportamento omogeneo da parte di tutte le persone è un obiettivo di primaria importanza.

Se si osservano le cose dal punto di vista del cliente tutto ciò diviene evidente. Solitamente, le interazioni del cliente con l'azienda sono gestite da una serie di dipendenti con ruoli ben precisi, in situazioni distinte. Non solo le persone coinvolte sono diverse, hanno competenze, esperienze e motivazioni differenti, ma spesso non sono al corrente delle altre interazioni che il cliente ha avuto. Fatto ancora peggiore, possono essere inconsapevoli di una strategia complessiva o di un livello di servizio auspicabile per gestire un determinato cliente. L'azienda può non vedere o capire le esperienze negative che sta procurando ai clienti, ma questi ultimi ne sono perfettamente consapevoli e agiranno in modo altrettanto negativo.

Dall'esperienza negativa all'abbandono dell'azienda, che inavvertitamente l'ha procurata, il passo del cliente è molto breve. Se ciò accade, il colpo accusato dall'azienda può essere molto duro; per qualsiasi azienda e in qualunque settore di attività è, infatti, molto meno costoso mantenere i clienti in essere che acquisirne di nuovi.

Per evitare che ciò accada l'azienda deve impegnarsi per comprendere le esigenze dei clienti, per affinare la propria offerta e attirare così nuovi clienti, oltre a mantenere quelli già acquisiti.

Per poter implementare in modo corretto questa filosofia sono necessari rilevanti investimenti tecnologici, organizzativi e commerciali per migliorare l'interazione con il cliente ed accrescere il valore dell'offerta dell'azienda nei suoi confronti; in altre parole è necessario implementare un sistema di CRM che consenta di gestire le relazioni con i clienti, mirando al loro mantenimento con strategie di fidelizzazione finalizzate ad incrementare la loro fiducia e la loro fedeltà nei confronti dell'azienda.

Un sistema di CRM è costituito da un insieme di strumenti in grado di tracciare la storia delle interazioni con i clienti, al fine di ottenere informazioni strategicamente importanti per il supporto alle attività di gestione del ciclo di vita del cliente.

Attraverso un sistema CRM l'azienda può identificare i propri clienti, riconoscendoli in ogni transazione indipendentemente dal canale di comunicazione, differenziare la gestione e personalizzare il rapporto con ciascuno di essi.

Sulla base delle informazioni che pervengono dal sistema di CRM l'azienda è in grado di definire con semplicità le strategie da utilizzare durante le fasi di selezione, acquisizione, mantenimento e crescita dei clienti al fine di raggiungere uno degli obiettivi principali del CRM ossia quello di facilitare il

processo di sviluppo della fedeltà e di conseguenza del livello di profittabilità, attraverso il ciclo di centratura sul cliente come mostra la figura 3.1.

I sistemi di CRM mirano ad incrementare il valore aziendale, facendo arrivare l'informazione giusta, alla persona giusta, al momento giusto lungo il canale che essa stessa ha scelto.



Figura 3.1 “Il ciclo di centratura sul cliente”

Per esser completo un sistema di CRM deve mettere a disposizione strumenti in grado di curare, tutti e tre, i seguenti aspetti¹⁸:

- operativo,
- analitico,
- collaborativo.

¹⁸ Greenberg P., *CRM Customer Relationship Management*, APOGEO, Milano, 2001.

L'aspetto operativo riguarda le funzioni aziendali tradizionali come l'assistenza alla clientela, gestione ordini, fatture/fatturazioni o automazione e gestione di vendite e marketing. In questo ambito il sistema deve porre il focus sull'efficienza operativa, fornendo strumenti di supporto ai processi operativi di marketing, vendite e servizi.

L'aspetto analitico di un sistema CRM è quello che cura la raccolta, la conservazione, l'estrapolazione, l'elaborazione, l'interpretazione e la comunicazione di informazioni sul cliente. Le aziende che forniscono questi strumenti hanno sviluppato applicazioni in grado di catturare questi dati da diverse fonti, conservarli nei database clienti e poi utilizzare centinaia di algoritmi per analizzarli/interpretarli in base alle richieste. Definendo il valore del programma applicativo in base alla capacità di fornire un servizio personalizzato.

L'aspetto collaborativo gestisce e coordina la comunicazione fra l'azienda e i clienti e fra azienda e fornitori, attraverso canali tradizionali come la posta o attraverso il web e le e-mail. Gli strumenti utilizzati a tal scopo possono essere un portale, un'applicazione per la gestione delle relazioni con i partner o un centro per l'interazione con i clienti. In altre parole, si tratta di qualsiasi strumento che fornisca supporto alla gestione della relazione con l'esterno (clienti, fornitori, partner).

Una volta definiti gli obiettivi e le strategie di CRM, la tecnologia può essere vista come un dispositivo in grado di metterli in atto. La tecnologia riveste un'importanza rilevante in ogni aspetto (operativo, analitico, collaborativo) di un sistema di CRM.

A livello operativo sono necessari strumenti informatici rivolti alla gestione delle attività di front-office come l'automazione della forza vendita e l'automazione del marketing.

A livello analitico è necessario l'impiego di strumenti in grado di archiviare i dati sui clienti, sezionarli e analizzarli per estrarre informazioni e conoscenza da presentare in forma vantaggiosa per l'azienda.

A livello collaborativo, servono strumenti di gestione dei punti di contatto con la clientela, ossia strumenti di comunicazione con i quali un cliente potrebbe interagire, fra i quali l'e-mail, le chiamate telefoniche, i fax, le pagine web.

Il “motore” del CRM è rappresentato dal database contenente le informazioni sui clienti¹⁹. Questo contenitore di informazioni costituisce il luogo dove tutti i dati sui clienti vengono immagazzinati e conservati. Potrebbe consistere in elementi basilari come il nome, l'indirizzo, il numero di telefono e la data di nascita, oppure contenere informazioni più complesse come il numero di accessi al sito, i movimenti fra le pagine e il tempo trascorso ad indugiare su di un prodotto. L'obiettivo è comunque quello di realizzare un unico punto di raccolta per le informazioni di ogni singolo cliente, in modo da ottenere una visione unitaria per tutti i reparti e le funzioni di un'azienda che hanno bisogno di conoscerle.

¹⁹ Raimondi M. et al., *CRM analitico: il contributo della business intelligence*, in “Sistemi & Impresa”, 9, 2003, pp. 75-77, ESTE, Milano.

3

Il caso “MasterDEA”

3.1

I servizi di formazione universitaria

Come introdotto nel capitolo 2 del presente lavoro, l’attenzione prestata nell’ultimo ventennio alle componenti ed alle determinanti del concetto di CRM, nonché all’elaborazione di tecniche e strumenti per la sua misurazione, risponde alla diffusa richiesta da parte dei clienti di una partecipazione e coinvolgimento nel processo che porta alla fornitura di beni e all’erogazione di servizi.

Il servizio rappresenta “un insieme di attività”, assistite o meno da prodotti, che nasce come risposta alle diverse esigenze del cliente. Maggiore sarà la capacità di conoscere e di comprendere i bisogni ed i desideri del cliente, maggiore sarà la probabilità di fornirgli le giuste risposte nei tempi e nei modi più idonei.

Una classificazione tradizionale suddivide i servizi nelle seguenti categorie¹.

- **Servizi di processo.** Si deve garantire la disponibilità di un bene attraverso il controllo di un processo (es. produzione di un qualunque bene materiale, energia elettrica, gas, ecc.).
- **Servizi di elaborazione.** Si produce un vero e proprio prodotto, come ad esempio, un software di elaborazione, un progetto, una ricerca ecc.; il valore del prodotto materiale è minimo, mentre alto è il valore intellettuale dell'opera (es. produzione di software, progetti di ricerca ecc.).
- **Servizi di disponibilità.** Il servizio consiste nel mettere a disposizione degli oggetti, come, ad esempio, automobili, alberghi, centri congressi ecc. (es. noleggio automezzi).
- **Servizi di interfaccia.** Vi è interazione tra fornitore e cliente, nella quale si esaurisce il servizio (es. servizi di formazione, sanitari, bancari, consulenziali, ecc.).

Tutti questi servizi rappresentano un'attività di problem solving per il cliente, ai quali l'erogatore dovrà essere in grado di rispondere fornendo beni e servizi adeguati. Tale capacità nasce da una comunicazione corretta con il cliente. A questo scopo è dedicata la pubblicità, i cui effetti sul cliente possono fornire all'azienda informazioni, che analizzate statisticamente, potrebbero essere utili a valutare l'efficacia del canale informativo utilizzato e a indirizzare i comportamenti di comunicazione aziendale. Il fine ultimo è, chiaramente, quello di raggiungere il maggior numero possibile di potenziali clienti cercando di instaurare con loro un rapporto di lungo termine dando la massima diffusione ai propri prodotti e servizi.

¹ D'Ambra L., Amenta P., *Un approccio statistico multidimensionale per la valutazione della performance dei servizi di interfaccia con informazioni esterne*, in "Qualità", vol. C, 2000, pp. 13-28, XX Convegno nazionale dell'AICQ, Bologna.

3.2

Il master: una scelta delicata in un mercato del lavoro incerto

Nel caso di studio, che si sta per affrontare, si sono presi in considerazione i dati forniti dall'Università di Pisa per quanto riguarda alcuni dei servizi di interfaccia, che fornisce attraverso il Dipartimento di Economia Aziendale: i Master DEA.

Prima di procedere, però, sarà utile chiarire cos'è un Master e a cosa serve.

Il **Master** è un titolo di studio superiore alla laurea. La sua prima funzione è quella di fornire un titolo di perfezionamento, accessibile solo dopo il conseguimento della laurea triennale o della laurea specialistica.

I Master si distinguono in *universitari* e *non universitari*. I primi sono soggetti a disposizioni di legge e sono parte integrante del sistema universitario, mentre i Master non universitari sono gestiti da enti privati che ne delineano autonomamente le caratteristiche. La differenza sostanziale tra i master universitari e corsi analoghi organizzati da altri enti consiste nel fatto che soltanto i primi fanno acquisire un titolo accademico legalmente riconosciuto.

I Master universitari, a seguito della recente riforma universitaria, si differenziano tra primo e secondo livello. Si accede ai primi a seguito di una laurea triennale, mentre per i Master di secondo livello è necessaria una laurea specialistica. I Master di primo livello presentano un ampio ventaglio di argomenti, prediligendo un aspetto multidisciplinare, mentre i Master di secondo livello sono rivolti alla massima specializzazione. Entrambi, però, devono prevedere, al loro termine, un periodo di stage obbligatorio.

I motivi che possono spingere alla scelta di frequentare un master possono essere i più diversi.

Il candidato che sceglie un Master non è essenzialmente un giovane neolaureato: le motivazioni che spingono a frequentare un Master sono molteplici e differenziate, così come i profili dei candidati che decidono di intraprendere questo percorso di specializzazione. Sempre più spesso, infatti, i master sono "post-experience", vale a dire che a frequentarli sono persone che hanno già avuto un'esperienza aziendale di 3-5 anni e che vogliono sistematizzare meglio quanto hanno già provato sul campo, consolidando ulteriormente la propria professionalità.

In linea generale però chi decide di continuare gli studi subito dopo la laurea può farlo perché vuole una professionalità specifica e spendibile, avvertendo come inadeguata dal punto di vista strettamente operativo la propria preparazione universitaria.

In generale la decisione di frequentare un Master verte sulla necessità di approfondire il proprio bagaglio formativo e acquisire competenze specialistiche fondamentali per l'orientamento professionale ricercato, sviluppando una vera e propria "identità lavorativa".

Per i neo-laureati può essere una scelta strategica importante per accedere concretamente al mondo del lavoro, mettendo in atto la formazione acquisita attraverso uno stage o ambire a posizioni di rilievo in organizzazioni dagli alti livelli di contenuti.

E' fondamentale considerare però che un Master post-laurea richiede un impegno e dei ritmi differenti da quelli universitari e lavorativi e un investimento economico che spesso risulta considerevole e va ad incidere sulla scelta finale. La decisione dovrà essere quindi accurata e in linea con il percorso ricercato.

Svolgere un master può essere importante anche per provare ad entrare nel mondo del lavoro grazie allo stage che completa questo tipo di proposta

educativa. Oggi, infatti, lavorare significa saper conciliare il bagaglio formativo accumulato e le ambizioni professionali ricercate. Tale obiettivo non è semplice da raggiungere, soprattutto in un mondo del lavoro che rende sempre più incerta la ricerca di un orientamento stabile ed impone un costante aggiornamento delle proprie competenze.

L'esperienza offerta da un Master acquisisce in questo contesto una valenza importante per il raggiungimento degli obiettivi professionali di ciascun candidato. Tuttavia la scelta di intraprendere o meno questa esperienza deve essere preceduta dall'approfondimento del significato che conserva e dalle conseguenze che comporta, al fine di non seguire un percorso spesso costoso sia in termini economici che di tempo e che, in ogni caso, raramente può dare garanzie di placement assolute.

La scelta di frequentare un master, quindi, è particolarmente carica di significato, oltre che di conseguenze. È perciò opportuno fare un po' di luce sull'argomento e cercare di capire se continuare il proprio percorso formativo ha senso oppure no.

Proprio in virtù di quanto finora esposto, stabilire quale master frequentare è una decisione decisamente delicata. I parametri da verificare nella scelta sono di diversi tipi: va di fatto valutata nel suo complesso la sua serietà e autorevolezza, prendendo in considerazione fattori di vario genere.

In primo luogo, ci si dovrà interrogare su quanto siano effettivamente richieste le professionalità che il master si ripromette di formare, sforzandosi di analizzare il mercato attuale. In questo senso, può essere utile concentrare i propri sforzi in tre direzioni:

- è fondamentale essere a conoscenza della serietà e prestigio dell'ente o dell'università che presenta il corso;

- va vagliato il piano didattico, capendone l'attuale portata e la sua "spendibilità";
- è necessario individuare le figure più apprezzate dalle aziende sfruttando i differenti canali a disposizione per cercare lavoro (siti Web di e-recruitment, quotidiani, periodici specializzati).

Vanno, inoltre, richieste agli organizzatori del master (e magari anche a chi l'ha già frequentato negli anni precedenti) informazioni dettagliate circa il placement che viene prospettato, cioè l'insieme delle attività che agevolano l'inserimento nel mondo del lavoro di chi ha frequentato il master.

Da ultimo, si devono considerare con attenzione i partner e i patrocinatori del progetto formativo, così da coglierne la credibilità e i reali collegamenti con il mondo del lavoro, non dimenticando di ponderare anche l'autorevolezza del comitato scientifico e del corpo docente.

3.3

I MasterDEA

Nell'Università di Pisa il Dipartimento di Economia Aziendale "E. Giannessi" è il centro a cui compete la ricerca rivolta a problematiche aziendali e condotta con metodologie economiche.

Costituito come unità organizzativa finalizzata alla promozione ed al coordinamento delle attività di ricerca nell'Università, in base al DPR. 382 del 1980, il Dipartimento è la struttura scientifica che riunisce settori di ricerca omogenei per fini o per metodo e i relativi insegnamenti, organizza le strutture

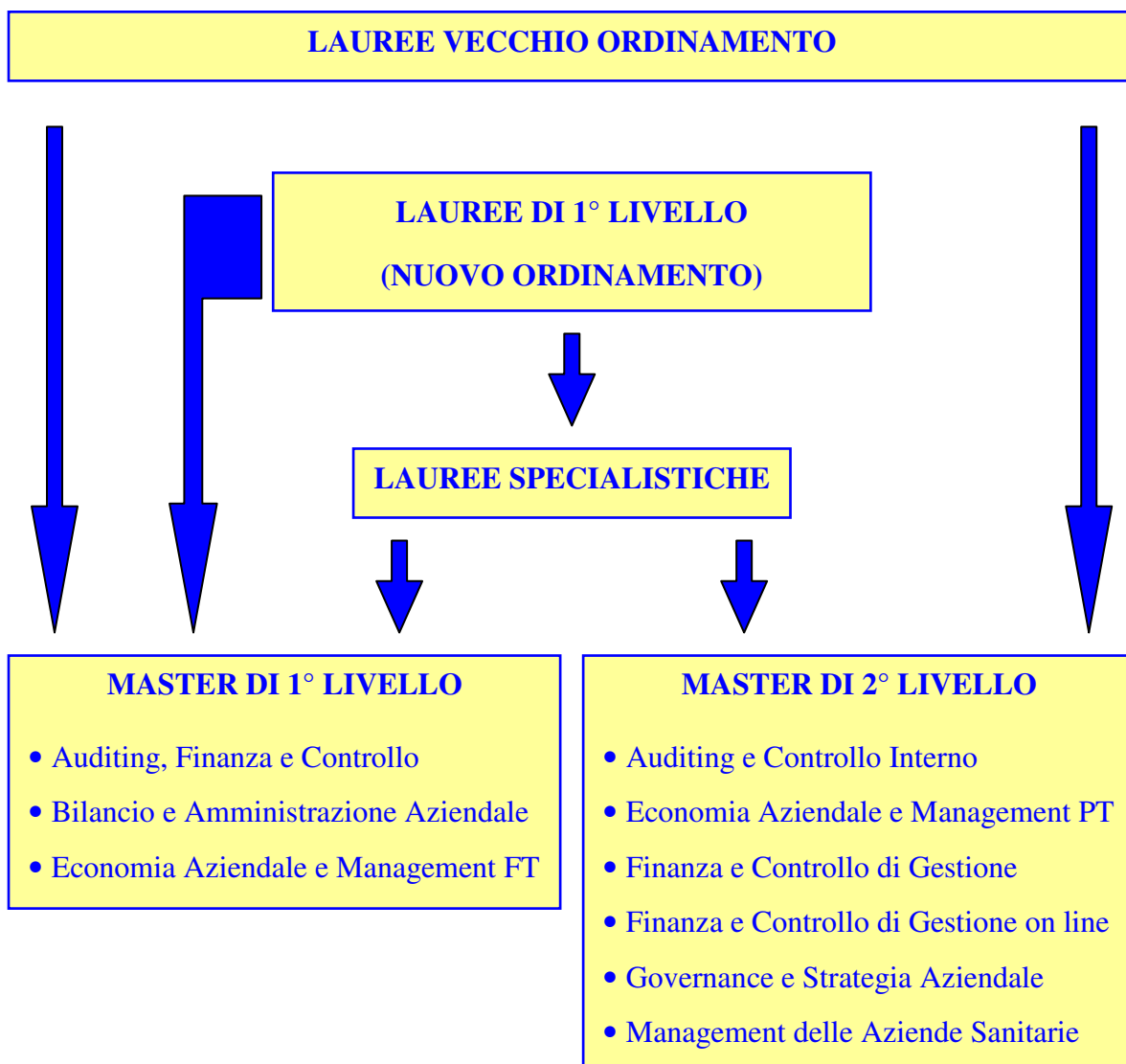
per la ricerca, è il soggetto a cui vengono affidati attività di consulenza e programmi di ricerca su contratto o convenzione.

Il dipartimento di economia aziendale da tempo ormai si propone quale centro per l'alta formazione di economisti aziendali che, attraverso una preparazione qualificata, potranno occupare posizioni di responsabilità nel mondo del lavoro, nell'industria, nei settori del credito, della finanza, nel commercio, nella pubblica amministrazione, nei centri di ricerca, negli organismi nazionali ed internazionali di programmazione economica. Questi obiettivi sono molti e molto ambiziosi. La risposta che il dipartimento di Economia Aziendale della Facoltà di Economia pisana intende dare, con l'attivazione della propria offerta formativa post-laurea (Formazione Avanzata Economia), mira a soddisfare le reali esigenze del mondo del lavoro, degli studi e della cultura economica-aziendale.

Con la sua offerta formativa, composta di master, seminari e workshop, la Formazione Avanzata Economia, si propone di:

- formare figure professionali in grado di potenziare lo sviluppo delle aziende industriali e di servizi;
- promuovere la diffusione della cultura manageriale ed imprenditoriale e favorire la nascita di nuove iniziative economiche;
- fornire una risposta adeguata alle esigenze formative emergenti nelle aziende private, pubbliche e nella libera professione.

L'offerta formativa post laurea MasterDEA è articolata secondo il seguente schema²:



² Brochure informativa MasterDEA, a.a. 2006/2007.

3.4

Il mercato della formazione

Negli ultimi anni, in seguito alle numerose riforme del sistema di istruzione italiano, il mercato della formazione ha subito numerosi cambiamenti. Il decreto n. 509/99 è quello che ha determinato il reale attuarsi della riforma universitaria: esso dà una serie di indicazioni generali per gli ordinamenti degli studi universitari. A partire da queste indicazioni i singoli atenei hanno preso autonome decisioni, definendo così, progressivamente, l'attuarsi di un sistema universitario effettivamente nuovo.

In questo contesto completamente stravolto da quelli che una volta erano i canoni dell'istruzione, i vari enti/aziende, universitari e non, devono continuamente difendere la propria immagine e conquistare nuove quote di mercato strappandole alla concorrenza con servizi di formazione che soddisfino le esigenze degli utenti finali. Per adattarsi a tali cambiamenti è necessario innanzitutto capire e conoscere i propri clienti. A questo proposito, anche per i servizi di formazione, sono diventate di fondamentale importanza le attività di marketing.

Il dipartimento di Economia Aziendale dell'Università di Pisa si è mosso in questa direzione ampliando la propria offerta formativa con un gli otto master post laurea sopraccitati, dedicando all'organizzazione MasterDEA risorse, umane e non, di qualità e implementando attività di marketing che fino a qualche anno fa erano considerate una competenza esclusiva delle imprese private.

Le attività di marketing MasterDEA sono attuate essenzialmente su due piani: la pubblicità attraverso i canali dei media e il contatto diretto con attività di *public relation*. La pubblicità è usata per promuovere l'offerta formativa, ma soprattutto è usata per diffondere e rafforzare l'idea di prestigio dell'ente organizzatore. Le attività di *public relation* sfruttano essenzialmente il contatto diretto e personale con i soggetti interessati al servizio e per la loro natura sono destinate a fornire quelle informazioni supplementari (prezzi, modalità di ammissione, servizi aggiuntivi) che difficilmente possono essere date attraverso i normali canali informativi.

Con il crescere del numero di master, e quindi del numero di allievi, hanno assunto fondamentale importanza nei processi decisionali la gestione dei dati e delle informazioni quantitative e il processo di informatizzazione che interessa soprattutto l'area del marketing.

L'obiettivo è quello di costruire un'infrastruttura tecnologica per reperire e memorizzare dati sui soggetti che sono stati raggiunti dalle attività pubblicitarie e che hanno presentato domanda di ammissione ad uno dei Master, al fine di costruire un ambiente di *datawarehousing* che permetta di fare delle analisi statistiche sui dati raccolti. I risultati di tali analisi servono a confermare l'utilità del progetto intrapreso nel suo complesso e a giustificarne la continuazione.

Il caso di studio qui analizzato riguarda il primo tentativo dell'area marketing MarsterDEA di analisi dei dati relativi alle campagne pubblicitarie effettuate per individuare quali, tra canali informativi utilizzati, sono i più efficaci; per cercare di individuare la migliore strategia di comunicazione; e per una migliore allocazione del budget pubblicitario.

3.5

Il caso

La prima fase dell'analisi ha riguardato la raccolta delle informazioni e dei dati sull'utente sia attraverso il recupero e l'integrazione di tutti i dati derivati dalla compilazione delle richieste di iscrizione, sia attraverso la raccolta dei dati delle frequenze di visita degli otto siti internet -uno per ciascun master- che sono stati pubblicati sul web.

Il responsabile delle attività di marketing dispone di un *database* composto da informazioni riguardanti i nominativi, con informazioni anagrafiche annesse, di coloro che presentano la domanda di ammissione. Queste informazioni, raccolte su moduli cartacei, vengono trascritte su supporti elettronici al fine di averne sempre una disponibilità immediata attraverso l'estrazione di report di sintesi.

Il presente caso di studio avrà come primo obiettivo quello di accedere a queste informazioni (spesso gelosamente custodite) analizzarle e integrarle con i risultati per creare un'unica base di dati condivisa e visibile da tutti gli uffici del Dipartimento di Economia Aziendale.

In secondo luogo si cercherà di individuare una valida modalità di utilizzo delle informazioni riguardanti le frequenze di visita da parte degli utenti del canale informativo internet che sono fornite a pagamento da società di servizi di rete come ad esempio "Google" con il suo servizio "Analytics".

3.5.1

Le informazioni

Come è stato detto nel precedente paragrafo, per la raccolta di informazioni e per la valutazione della pubblicità MasterDEA è stato preso in considerazione un campione con numerosità pari a 347 estratto casualmente tra tutte le domande di ammissione pervenute alla segreteria MasterDEA per l'iscrizione ai corsi dell'anno accademico 2006/2007. Di ciascuna domanda è stata presa in considerazione la sola parte relativa ai canali informativi, attraverso i quali il candidato è venuto a conoscenza dei master. Si sono tralasciate le parti relative ai dati anagrafici, al percorso formativo, alle esperienze lavorative pregresse, poiché poco rilevanti ai fini dell'analisi qui condotta (tab. 3.1).

CANALI INFORMATIVI	BAA	EAM	FCG	MANSAN	GSA	ACI	AFC	FCG_OL	TOT
RICERCA ON LINE	4	14	16	6	3	44	21	22	130
STAMPA	1	2	2	3	5	7	3	37	60
SITO WEB	5	3	11	7	6	13	4	2	51
BROCHURE	8	3		4		12	1	1	29
AMICI	2	1	3	5	1	11		2	25
AFFISSIONI	1	2	1			4	5		13
ENTI	1		1		3	3	2	1	11
E-MAIL	1	2	1	2		3		1	10
PROFESSORI	2	3	2					1	8
MEETING	1	1			2	1			5
COMUNICAZIONE POSTALE		1				2			3
RADIO								2	2
TOT	26	32	37	27	20	100	36	69	347

Tabella 3.1. Canali informativi distinti per tipologia di master.

È evidente, però, che se si facesse un'analisi della tabella precedente, si potrebbe giungere a risultati poco significativi, data la scarsa frequenza rilevata per alcuni canali informativi. Di conseguenza si è deciso di raggruppare, in tre macro classi, i dati relativi ai canali informativi assimilabili (tab. 3.2).

MASTER	Internet	Stampa	Altro	TOTALE
BAA	10	10	6	26
EAM	19	8	5	32
FCG	28	3	6	37
MANSAN	15	7	5	27
GSA	9	5	6	20
ACI	60	25	15	100
AFC	25	9	2	36
FCG_OL	25	38	6	69
TOTALE	191	105	51	347

Tabella 3.2. Sintesi canali informativi.

Le tre macro-classi individuate sono:

- *internet*, che sintetizza i canali “ricerca on line”, “sito web” ed “e-mail”;
- *stampa*, dove convergono i dati relativi alla “stampa”, “brochure”, “affissioni” e “comunicazione postale”;
- *altro*, che sintetizza le voci “amici”, “enti”, “professori”, “meeting”, “radio”.

Entrando nel merito delle singole macro-classi è possibile evidenziare, in termini di frequenze assolute, che internet è senz'altro il mezzo di comunicazione più utilizzato da coloro che cercano informazioni sull'offerta formativa post-laurea. In alcuni casi (vedi master FCG_OL), però, la “stampa” riveste un ruolo di primo piano nella comunicazione, mentre la voce “altro” in

tutti i casi riveste un ruolo marginale e forse di supporto alla ricerca di informazioni on-line. Queste, tuttavia, sono solo delle supposizioni che dovranno essere verificate con appropriate analisi statistiche.

Altre informazioni utili ai fini dell'analisi del presente caso di studio sono le frequenze di visita dei siti internet che sono stati pubblicati per promuovere l'offerta formativa MasterDEA. Tali dati sono stati estratti dal database di Google e raggruppati per frequenze mensili (tab. 3.3). Il periodo di riferimento preso in considerazione è di dodici mesi (dal 1/2/2006 al 31/1/2007).

	Master BAA	Master EAM	Master ACI	Master FCG	Master FCG_OL	Master GSA	Master MANSAN	Master AFC	TOT
MESI									
feb.	2076	1695	3385	1589		2216	1159		12120
mar.	1337	3176	4386	1351		1431	1050		12731
apr.	1049	2817	2864	970		932	820		9452
mag.	953	1328	3562	881	967	978	749		9418
giu.	409	1150	4280	1138	8678	982	825	168	17630
lug.	969	1059	5506	1140	3509	941	729	992	14845
ago.	738	874	3874	836	2026	745	571	682	10346
sett.	643	1322	8031	1089	5897	1966	1376	1714	22038
ott.	1266	1418	7282	1151	1143	3713	2036	657	18666
nov.	1054	1941	4158	1546	938	1220	732	708	12297
dic.	1666	2206	4168	2025	1481	1168	835	758	14302
gen.	2886	3444	4720	3853	1115	1431	899	980	19238
TOT	15046	22425	56216	17569	25754	17723	11781	6659	173173

Tabella 3.3. Frequenze assolute di visite on line.

Per ognuna di queste frequenze mensili è stata calcolata una media giornaliera al fine di analizzare il trend delle visite e vedere come variano nell'arco di tempo considerato (tab. 3.4).

	Master BAA	Master EAM	Master ACI	Master FCG	Master FCG_OL	Master GSA	Master MANSAN	Master AFC
MESI								
feb.	74,143	60,536	120,893	56,750		79,143	41,393	
mar.	43,129	102,452	141,484	43,581		46,161	33,871	
apr.	34,967	93,900	95,467	32,333		31,067	27,333	
mag.	30,742	42,839	114,903	28,419	34,536	31,548	24,161	
giu.	31,462	38,333	142,667	37,933	289,267	32,733	27,500	33,6
lug.	31,258	34,161	177,613	36,774	113,194	30,355	23,516	32
ago.	23,806	28,194	124,968	26,968	65,355	24,032	18,419	22
sett.	32,150	44,067	267,700	36,300	196,567	65,533	45,867	57,133
ott.	40,839	45,742	234,903	37,129	36,871	119,774	65,677	21,194
nov.	35,133	64,700	138,600	51,533	31,267	40,667	24,400	23,6
dic.	53,742	71,000	134,452	65,323	47,774	37,667	26,935	24,452
gen.	93,097	111,097	152,258	124,290	35,968	46,161	29,000	31,613

Tabella 3.4. Numero medio giornaliero di accessi ai siti MasterDEA.

Un'analisi del trend del numero delle connessioni ai siti MasterDEA può essere molto utile per verificare se la campagna pubblicitaria ha avuto degli effetti sui comportamenti dei soggetti che sono alla ricerca di informazioni sulle offerte formative post-laurea; a questo scopo è stato utile riportare i valori medi della tabella precedente su un sistema di assi cartesiani in modo da mettere in evidenza eventuali variazioni degli accessi giornalieri (fig. 3.5).

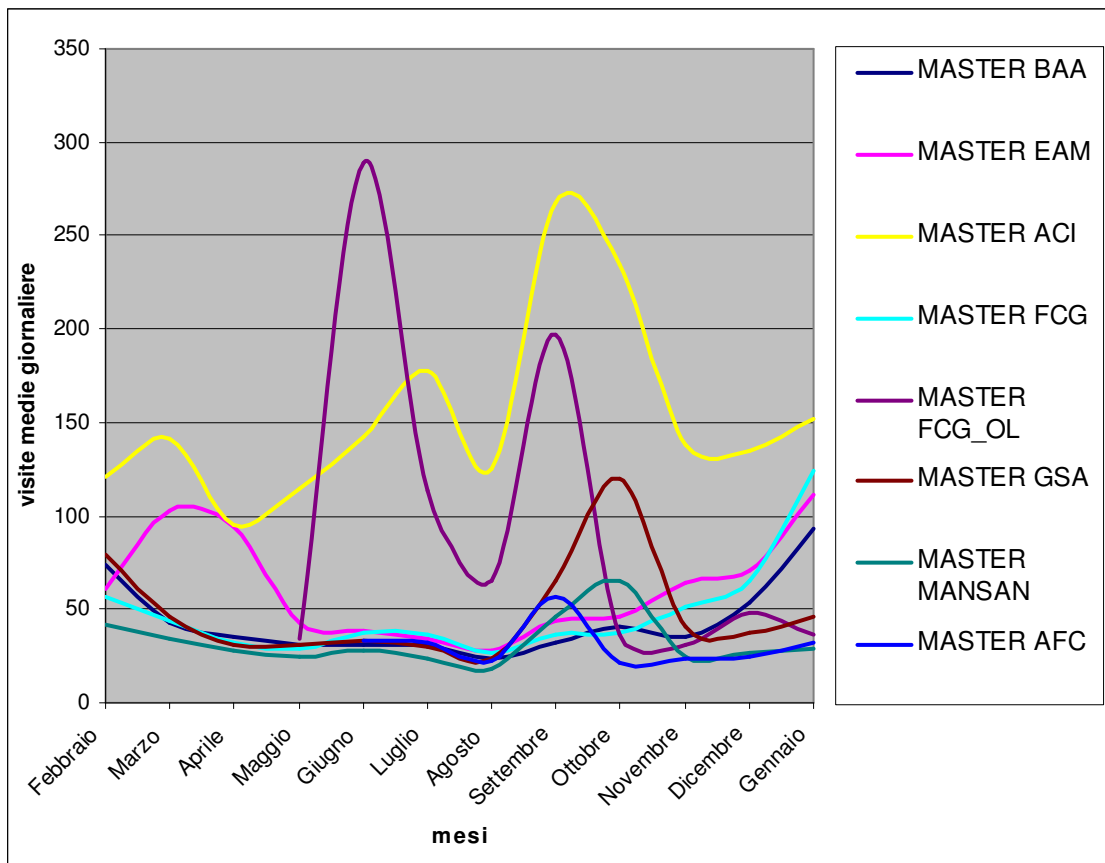


Figura 3.5. Trend degli accessi ai siti MasterDEA

Come si può notare, il numero medio delle connessioni giornaliere cresce notevolmente in alcuni periodi dell'anno. Se tali periodi, vengono confrontati con i periodi di campagna pubblicitaria, è possibile riscontrare una certa coincidenza. Infatti per l'anno accademico 2006/2007 le campagne pubblicitarie, per ciascun master, hanno interessato i seguenti periodi:

- Master BAA: dal 20/11/2006 al 05/02/2007;
- Master EAM: dal 07/11/2006 al 20/02/2007;
- Master ACI: dal 01/06/2006 al 10/10/2006;
- Master FCG: dal 06/11/2006 al 31/01/07;
- Master FCG_OL: dal 09/05/2006 al 30/09/2006;
- Master GSA: dal 01/09/2006 al 28/02/2007;

- Master MANSAN: dal 12/09/2006 al 09/10/2006;
- Master AFC: dal 29/06/2006 al 10/10/2006.

Tale coincidenza induce a pensare che le campagne pubblicitarie MasterDEA influenzano effettivamente il comportamento di colui che, intenzionato ad accrescere le proprie competenze specialistiche con un percorso formativo post-laurea, si connette ad internet per ricercare il maggior numero possibile di informazioni su ciò che il mercato della formazione offre.

3.6

L'analisi statistica dei dati

Una volta individuate le possibili variazioni nei comportamenti di visita dei siti MasterDEA sarà necessario verificare, attraverso analisi statistiche, la conformità, la coerenza, e l'usabilità delle informazioni rispetto ai fatti che si vogliono analizzare.

La verifica di *conformità* consiste nel confronto tra i fatti analizzati e l'universo considerato. Una volta esplorato l'universo dei dati e prodotte una serie di fotografie, più o meno dettagliate e articolate sui comportamenti di risposta agli input pubblicitari, è necessario analizzare il comportamento di colui che presenta domanda di iscrizione attraverso la formulazione di ipotesi statistiche.

Il concetto di *coerenza*, fa proprio riferimento al confronto tra le ipotesi formulate, riguardanti la popolazione di riferimento, e la realtà indagata. Per giungere ad un test delle ipotesi, è necessario tradurre un'affermazione del

mondo reale in un'affermazione concernente la distribuzione di probabilità di una variabile casuale, la cui veridicità si intende controllare sulla base delle evidenze campionarie.

Una volta chiarita la situazione sia dal punto di vista dei dati disponibili, delle tecnologie e degli strumenti da usare, sarà necessario valutare l'*usabilità* dei risultati ottenuti. L'analista dovrà comunicare al responsabile marketing le possibili vie da percorrere per migliorare la comunicazione pubblicitaria, ridefinendo l'importanza: di un sistema di *data warehouse*; delle aspettative connesse al tipo di informazioni restituite dall'analisi; e sull'utilità di queste ultime.

Lo scopo di questo caso è presentare un'analisi dei dati di visita a un sito web con due obiettivi principali e parzialmente distinti. Nella prima parte del caso si cercherà di estrarre informazioni preziose sui comportamenti di visita. Nella seconda parte l'obiettivo sarà quello di esprimere un giudizio sulla validità delle scelte di comunicazione MasterDEA.

Per raggiungere questi obiettivi, in primo luogo saranno condotte analisi statistiche finalizzate alla descrizione delle variazioni dei comportamenti di visita tra il "prima" e il "dopo" la campagna pubblicitaria. Di fondamentale importanza sarà anche il confronto tra i vari mezzi di comunicazione utilizzati al fine di valutarne l'efficacia. Nella seconda parte l'obiettivo sarà stimare i costi sostenuti rapportandoli ai benefici conseguiti. L'analisi sarà, quindi, indirizzata verso la valutazione di efficienza del canale o dei canali informativi più rilevanti in termini di influenza su colui che è interessato a presentare la sua candidatura per la frequenza ad uno dei MasterDEA.

3.6.1

L'analisi “prima - dopo”

Per misurare la relazione tra la pressione pubblicitaria e la variazione dei comportamenti di accesso ai siti MasterDEA, che nel periodo considerato ha effettuato consistenti investimenti pubblicitari, è stato utilizzato l'indice di eterogeneità di Gini che è espressione della mutabilità del fenomeno osservato.

Indice assoluto

$$G = 1 - \sum_i f_i^2$$

Indice normalizzato

$$G^* = \frac{k G}{k - 1}$$

dove:

f_i = frequenze relative;

$i = 1, 2, \dots, k$.

Per una variabile qualitativa (come quella indagata) la mutabilità esprime la possibilità di variare tra una perfetta omogeneità e una completa eterogeneità.

La minima eterogeneità (ovvero perfetta omogeneità) si verifica allorquando tutte le frequenze relative sono accentrate in un solo attributo: in tale situazione, ci attendiamo che l'indice normalizzato di eterogeneità valga zero ($G^* = 0$).

La massima eterogeneità (ovvero totale disomogeneità) si verifica quando le frequenze relative si distribuiscono in parti esattamente uguali tra più attributi: in tale situazione, ci attendiamo che l'indice normalizzato di eterogeneità raggiunga il massimo valore, cioè l'unità ($G^* = 1$).

Nei casi reali vi sarà una maggiore o minore accentuazione di un attributo rispetto agli altri, per cui il confronto tra le distribuzioni avviene mediante misure che fanno riferimento agli estremi di massima o minima omogeneità ovvero, di minima o massima eterogeneità³.

Ai fini dell'analisi è stato necessario prendere in considerazione le frequenze assolute di visita ai siti internet, ripartite in due periodi: uno precedente ed uno successivo all'azione pubblicitaria. Dato che i master considerati hanno delle scadenze diverse e quindi dei tempi pubblicitari diversi, come periodi di riferimento sono stati presi i quattro mesi precedenti e i quattro mesi successivi alla data di inizio della campagna di comunicazione (tab. 3.6 e tab. 3.7). Come è facile notare i MasterDEA qui presi in considerazione sono solo sei. I master FCG_OL e AFC non possono essere considerati perché, essendo alla loro prima edizione, non si hanno a disposizione le frequenze di accesso ai relativi siti web prima della campagna pubblicitaria.

BAA		EAM		ACI		FCG		GSA		MANSAN		TOT
MESI		MESI		MESI		MESI		MESI		MESI		
giu.	409	giu.	1150	feb.	3385	giu.	1138	giu.	982	mar.	1050	
lug.	969	lug.	1059	mar.	4386	lug.	1140	lug.	941	apr.	820	
ago.	738	ago.	874	apr.	2864	ago.	836	ago.	745	mag.	749	
sett.	643	sett.	1322	mag.	3562	sett.	1089	sett.	1966	giu.	825	
TOT	2759		4405		14197		4203		4634		3444	33642

Tabella 3.6. Frequenze assolute di visita prima della pubblicità.

BAA		EAM		ACI		FCG		GSA		MANSAN		TOT
MESI		MESI		MESI		MESI		MESI		MESI		
ott.	1266	ott.	1418	giu.	4280	ott.	1151	ott.	3713	lug.	729	
nov.	1054	nov.	1941	lug.	5506	nov.	1546	nov.	1220	ago.	571	
dic.	1666	dic.	2201	ago.	3874	dic.	2025	dic.	1168	sett.	1376	
gen.	2886	gen.	3444	sett.	8031	gen.	3853	gen.	1431	ott.	2036	
TOT	6872		9004		21691		8575		7532		4712	58386

Tabella 3.7. Frequenze di visita assolute dopo la pubblicità.

³ Piccolo D., *Statistica per le decisioni*, Il Mulino, Bologna, 2004.

Dato che l'eterogeneità misura la variabilità delle frequenze relative, queste ultime potranno essere calcolate rapportando le singole frequenze mensili al totale del periodo considerato. Le frequenze relative così ottenute sono sintetizzate nella tabella seguente (tab. 3.8).

	prima della pubblicità	dopo la pubblicità
MASTER		
BAA	0,082	0,118
EAM	0,131	0,154
ACI	0,422	0,372
FCG	0,125	0,147
GSA	0,138	0,129
MANSAN	0,102	0,081
Totali	1,000	1,000

Tabella 3.8. Frequenze di visita relative prima-dopo la pubblicità.

Mediante tali informazioni, il calcolo dell'indice di eterogeneità di Gini è immediato:

$$G_{\text{prima}} = 1 - (0.082^2 + 0.131^2 + \dots + 0.102^2) = 0.753$$

$$G_{\text{dopo}} = 1 - (0.118^2 + 0.154^2 + \dots + 0.081^2) = 0.780$$

Da questi indici, essendo il numero dei master considerati pari a $k = 6$, si deducono le corrispondenti misure normalizzate.

$$\left. \begin{array}{l} G^*_{\text{prima}} = 0.904 \\ G^*_{\text{dopo}} = 0.936 \end{array} \right\} \text{Maggiore eterogeneità}$$

Questi risultati mostrano che prima della campagna pubblicitaria le visite on line tra i vari master sono più omogenee, nel senso che i siti vengono visitati in modo indifferente da soggetti in cerca di informazioni generali. La pubblicità ha, invece, l'effetto di indirizzare i visitatori verso i siti internet di loro interesse facendone aumentare le visite e rendendo più eterogenei i comportamenti di visita soprattutto nei periodi antecedenti la data di scadenza in cui ricadono le maggiori attività pubblicitarie.

Questo evidenzia il fatto che la pubblicità ha delle influenze notevoli sui comportamenti di coloro che sono alla ricerca di informazioni sull'offerta formativa post-laurea; di conseguenza, ai fini dell'analisi qui condotta sarà opportuno distinguere il canale informativo più efficace, tra quelli utilizzati.

3.6.2

Analisi di connessione tra i canali informativi

Per giungere all'individuazione del canale informativo più efficace è necessario indagare sull'eventuale esistenza di connessioni tra le tre macroclassi precedentemente individuate: "internet", "stampa", "altro". Per evidenziare queste relazioni si può ricorrere a misure di connessione che utilizzano le sole frequenze attraverso cui un fenomeno si manifesta. Con tali indicatori ciò che si vuole analizzare, in una distribuzione multivariata, è l'informazione connessa al contemporaneo presentarsi delle modalità di una variabile con quelle di un'altra variabile.

Una misura adeguata a misurare la dipendenza tra variabili è l'indice quadratico di connessione di Karl Pearson.

$$\chi^2_{(r-1)(c-1)} = \sum_i \frac{(f_{oi} - f_{ti})^2}{f_{ti}}$$

dove:

$r = n^\circ \text{ righe}$
 $c = n^\circ \text{ colonne}$

} per il calcolo dei gradi di libertà;

f_{oi} = frequenze osservate;

f_{ti} = frequenze teoriche attese.

Il problema statistico è quello di verificare quale delle seguenti affermazioni (ipotesi statistiche) sia sostenibile sulla base dell'evidenza sperimentale:

- H_0 : le due variabili categoriche sono indipendenti;
- H_1 : le due variabili categoriche sono dipendenti.

L'indice quadratico di connessione è proprio un indicatore che ci permetterà di accettare o rifiutare l'ipotesi H_0 di indipendenza tra i canali informativi considerati.

Per il calcolo dell'indice di connessione è necessario calcolare preliminarmente le frequenze teoriche attese, partendo dalle frequenze osservate, attraverso la seguente formula:

$$f_{ti} = \frac{n_{i.} * n_{.j}}{N}$$

indicando con:

$n_{i.}$ le frequenze totali di riga;

$n_{.j}$ frequenze totali di colonna;

N la numerosità del campione considerato.

La tabella seguente (tab. 3.9) mostra le frequenze osservate e (tra parentesi) le corrispondenti frequenze attese calcolate con la precedente espressione (f_{ii}).

MASTER	Internet	Stampa	Altro	TOTALE
BAA	10 (14,31)	10 (7,87)	6 (3,82)	26
EAM	19 (17,61)	8 (9,68)	5 (4,70)	32
FCG	28 (20,37)	3 (11,20)	6 (5,44)	37
MANSAN	15 (14,86)	7 (8,17)	5 (3,97)	27
GSA	9 (11,01)	5 (6,05)	6 (2,94)	20
ACI	60 (55,04)	25 (30,26)	15 (14,70)	100
AFC	25 (19,82)	9 (10,89)	2 (5,29)	36
FCG_OL	25 (37,98)	38 (20,88)	6 (10,14)	69
TOTALE	191	105	51	347

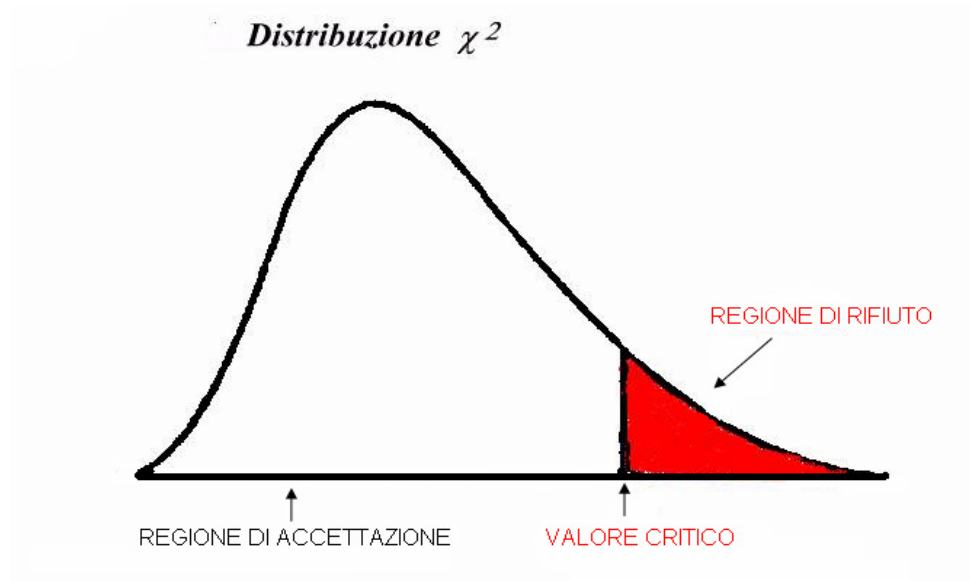
Tabella 3.9. Frequenze assolute e attese.

Con tutti i dati a disposizione il calcolo dell'indice quadratico di connessione χ^2 è immediato:

$$\chi^2 = \frac{(10 - 14,31)^2}{55,04} + \frac{(19 - 17,61)^2}{19,61} + \dots + \frac{(2 - 5,29)^2}{5,29} + \frac{(6 - 10,14)^2}{10,14} = 41,90$$

Confrontando il χ^2 calcolato con il con il χ^2 teorico con 14 gradi di libertà e con livello di significatività $\alpha = 0,01$ e $\alpha = 0,05$

$$\left. \begin{array}{l} \chi^2_{(14; 0.01)} = 29,141 \\ \chi^2_{(14; 0.05)} = 23,685 \end{array} \right\} \text{valori critici} \quad \chi^2 = 41,90 \quad \text{valore calcolato}$$



risulta che in entrambi i casi il χ^2 calcolato $> \chi^2$ teorico, quindi vi è sufficiente evidenza per rifiutare l'ipotesi di indipendenza visto che ci troviamo per entrambi i livelli di α nella regione di rifiuto.

Tali risultati mostrano che c'è una connessione tra le macro-classi considerate. Esiste, dunque, un tipo di pubblicità che non può prescindere dalle altre perché tutte concorrono a far conoscere ai soggetti interessati l'offerta formativa post-laurea MasterDEA. Fermo restando la validità di tutti i canali informativi utilizzati, l'individuazione del mezzo di comunicazione più efficace resterà un valido obiettivo di questa analisi per giungere ad una migliore ripartizione del budget pubblicitario.

3.6.3

L'ANOVA a due criteri di classificazione: canale informativo – master

Verificare l'esistenza di differenze significative tra le medie del campione considerato attraverso l'Analisi della Varianza è il passo successivo. Nella pratica sperimentale è molto comune il caso in cui debbano essere messe a confronto simultaneamente più di due popolazioni, cioè valutare nello stesso esperimento più di due campioni ai quali si riferiscono le misurazioni della variabile esaminata. Anche in questo caso, per vedere se esistono differenze significative tra i canali informativi e tra le otto tipologie di master sarà utilizzata un'ANOVA a due criteri di classificazione. Con tale analisi si cercherà di capire se ad influenzare il comportamento di iscrizione ai MasterDEA concorra la tipologia di master o il mezzo di comunicazione utilizzato.

Se indichiamo con a il fattore “tipologia master” e con b il fattore “canale informativo”, il modello dell'ANOVA può essere così formulato:

$$y_{ij} = \mu + \alpha_i + \beta_j + \varepsilon_{ij}$$

Le ipotesi statistiche che dovranno essere testate sono:

- H_0 : $\alpha_i = 0$ per ogni i
 $\beta_j = 0$ per ogni j
- H_1 : almeno un $\alpha_i \neq 0$ per qualche i
almeno un $\beta_j \neq 0$ per qualche j

La variabilità dei dati appartenenti ad un campione può essere rappresentata come scarto dalla media generale, cioè come *devianza*. A sua volta la devianza può essere scomposta in due parti: lo scostamento di ciascun dato dalla media del gruppo a cui appartiene (devianza interna) e quello della media stessa del gruppo dalla media generale (devianza esterna). Prendendo in considerazione le frequenze osservate della tabella 3.2 (qui di seguito riportata) a cui sono state aggiunte le medie di riga, di colonna e la media totale,

MASTER	Internet	Stampa	Altro	TOTALE	Medie
BAA	10	10	6	26	8,667
EAM	19	8	5	32	10,667
FCG	28	3	6	37	12,333
MANSAN	15	7	5	27	9
GSA	9	5	6	20	6,667
ACI	60	25	15	100	33,333
AFC	25	9	2	36	12
FCG_OL	25	38	6	69	23
TOTALE	191	105	51	347	
Medie	23.875	13.125	6.375		14.458

Media totale

Tabella 3.2. Sintesi canali informativi.

è possibile calcolare:

- la devianza totale;

$$SQ(t) = \sum_i \sum_j (\bar{y}_{ij} - \bar{y})^2 = 4203,958$$

- la devianza esterna attribuibile al fattore a;

$$SQ(a) = \sum_i (\bar{y}_{i.} - \bar{y})^2 = 1734,557$$

- la devianza esterna attribuibile al fattore b;

$$SQ(b) = \sum_j (\bar{y}_{.j} - \bar{y})^2 = 1246,333$$

- la devianza interna ovvero d'errore.

$$SQ(e) = SQ(t) - SQ(a) - SQ(b) = 1223,068$$

Il rapporto tra le devianze e i rispettivi gradi di libertà forniscono le stime della varianza.

I risultati dell'ANOVA possono essere riassunti come segue:

Fonte di variazione	Devianza (SQ)	gradi di libertà	Varianze (MQ)
Fattore a (master)	1734,557	7	247,794
fattore b (pubblicità)	1246,333	2	623,167
fattore e	1223,068	14	87,362
TOTALE	4203,958	23	

Le ipotesi H_0 per essere accettate o respinte devono essere saggiate con il test F. I valori empirici di F possono essere calcolati rapportando le varianze dei due fattori considerati (master e pubblicità) alla varianza d'errore, quindi:

$$F(a) = MQ(a) / MQ(e) = 2,836$$

$$F(b) = MQ(b) / MQ(e) = 7,133$$

Confrontando questi valori empirici con quanto atteso dalla distribuzione di F individuata dai corrispondenti gradi di libertà, si verifica se il rapporto è compatibile con H_0 in funzione del margine d'errore (livello di significatività) concesso.

Con un livello di significatività pari a 0,05 i valori teorici di F sono:

$$F_{(0,05; 7; 14)} = 2,77$$

$$F_{(0,05; 2; 14)} = 3,74$$

Dal confronto di tali valori con gli F empirici risulta che:

$$F(a) > F_{(0,05; 7; 14)} \quad \text{rifiuto l'ipotesi } H_0$$

$$F(b) > F_{(0,05; 2; 14)} \quad \text{rifiuto l'ipotesi } H_0$$

Dai risultati dell'ANOVA, con un livello di significatività pari a 0,05, è possibile dedurre che esistono delle differenze tra i diversi master (fattore *a*) e tra i diversi canali informativi utilizzati (fattore *b*). Quindi la tipologia di master e il canale utilizzato influenzano il comportamento di coloro che presentano domanda di iscrizione ai master.

Se il margine d'errore viene ridotto dal 5% all'1% risulta che:

$$F_{(0,01; 7; 14)} = 4,28$$

$$F_{(0,01; 2; 14)} = 6,51$$

Ne consegue che:

$$F(a) < F_{(0,01; 7; 14)} \quad \text{accetto l'ipotesi } H_0$$

$$F(b) > F_{(0,01; 2; 14)} \quad \text{rifiuto l'ipotesi } H_0$$

Con un livello di significatività pari a 0,01, dunque, i risultati dell'ANOVA consentono di affermare che esistono delle differenze tra le medie dei canali informativi, ma che le differenze tra un master ed un altro non sono significative. In questo secondo caso l'unica variabile in grado di influenzare il comportamento di iscrizione è la pubblicità.

A questo punto può essere interessante individuare quale tra le medie dei canali informativi utilizzati si discosta dalle altre, facendo due confronti: uno per testare la significatività della media della macroclasse “internet” dalle altre due, e uno per saggiare la differenza tra i gruppi “stampa” ed “altro”.

Il metodo maggiormente utilizzato nella verifica delle differenze tra le medie è quello dei “contrasti ortogonali”. In modo compatto e generico, un contrasto può essere scritto nella forma:

$$l(c) = \sum_i c_i \bar{x}_i$$

ove \bar{x}_i è la media campionaria del trattamento i-esimo, con varianza

$$MQ(l) = \sum_i c_i^2 \frac{\sigma^2}{n_i} = \sigma^2 \sum_i \frac{c_i^2}{n_i}$$

Il contrasto è definito dai valori che vengono assegnati ai coefficienti c_i , per comodità generalmente numeri interi. Se si considerano le differenze tra le medie, i contrasti sono caratterizzati dalla seguente condizione:

$$\sum_i c_i = 0$$

Applicando il metodo dei contrasti ortogonali ai risultati dell'ANOVA con un livello di significatività pari a 0,01, i confronti pianificabili prima di ottenere i risultati sono:

$$l_1 = 2 * 23,875 - 13,125 - 6,375 = 28,25$$

$$l_2 = 13,125 - 6,375 = 6,75$$

I due contrasti permettono di verificare due ipotesi H_0 :

- l_1 che non vi sia differenza tra le media di “internet” e quelle degli altri due canali informativi;
- l_2 che non vi siano differenze tra le medie di “stampa” e “altro”.

Anche in questo caso, come per l'ANOVA, la significatività dei contrasti deve essere stimata con un test F. Si procede, quindi, con il calcolo degli F empirici che dovranno poi essere confrontati con gli F teorici.

Per l_1 :

$$F_1 = MQ(l_1) / MQ(e) = (28,25^2 / (4/8 + 1/8 + 1/8)) / 87,362 = 12,18$$

Per l_2 :

$$F_2 = MQ(l_2) / MQ(e) = (6,752 / (1/8 + 1/8)) / 87,362 = 2,086$$

Dal confronto di tali rapporti con $F_{(0,01; 1; 14)} = 8,86$ risulta che:

$$\begin{array}{ll} F_1 > F_{(0,01; 1; 14)} & \text{rifiuto l'ipotesi } H_0 \\ F_2 < F_{(0,01; 1; 14)} & \text{accetto l'ipotesi } H_0 \end{array}$$

Dal primo confronto risulta una differenza significativa tra la media del canale internet e le altre due medie. Dal secondo confronto, invece, non risultano differenze significative tra la stampa e gli altri canali informativi utilizzati. E', quindi, possibile concludere, con un margine d'errore pari all'1%, che “internet”, rispetto alla “stampa” e ad “altri” canali informativi, risulta essere il mezzo pubblicitario più efficace.

3.6.4

Le differenze tra i master di primo e secondo livello

Con l'ANOVA a due criteri di classificazione siamo riusciti ad individuare quali tra i canali informativi utilizzati risulta essere più efficace nella comunicazione con coloro che sono alla ricerca di informazioni sull'offerta formativa post-laurea. Come anticipato all'inizio di questo capitolo, è possibile fare una distinzione tra master di primo e di secondo livello. È facile intuire che queste due categorie soddisfino esigenze diverse e siano rivolti a soggetti diversi. Di conseguenza, ai fini dell'analisi condotta potrebbe essere utile approfondire, con indagini statistiche simili a quelle del paragrafo precedente, la convenienza ad utilizzare mezzi di comunicazione differenti per ciascun master.

Prima di procedere con il calcolo statistico è necessario distinguere il campione di soggetti finora utilizzato indistintamente (vedi tab. 3.2), in due sottogruppi: uno per i master di primo livello (tab. 3.10), l'altro per quelli di secondo livello (tab. 3.11).

Master di I livello	Internet	Stampa	Altro	TOTALE	Medie
AFC	25	9	2	36	12
BAA	10	10	6	26	8,667
EAM Full Time	10	6	3	19	6,333
TOTALE	45	25	11	81	
Medie	15	8,333	3,667		9

Media Totale

Tabella 3.10. Canali informativi distinti per i master di I livello.

Master II livello	Internet	Stampa	Altro	TOTALE	Medie
ACI	60	25	15	100	33,333
EAM Part Time	9	2	2	13	4,333
FCG	28	3	6	37	12,333
FCG_OL	25	38	6	69	23
GSA	9	5	6	20	6,667
MANSAN	15	7	5	27	9
TOTALE	146	80	40	266	
Medie	24,333	13,333	6,667		14,778

Media Totale

Tabella 3.11. Canali informativi distinti per master di II livello.

Il campione, così distinto, dovrà essere analizzato con un'ANOVA a due criteri di classificazione da applicare, in modo totalmente indipendente, prima all'uno e poi all'altro sottogruppo.

Le ipotesi nulle ed alternative che anche in questo caso dovranno essere verificate sono:

- H_0 : $\alpha_i = 0$ per ogni i
 $\beta_j = 0$ per ogni j
- H_1 : almeno un $\alpha_i \neq 0$ per qualche i
almeno un $\beta_j \neq 0$ per qualche j

Master di I livello

Considerando le frequenze riscontrate per i master di primo livello (tab. 3.10) è necessario procedere con il calcolo delle devianze (esterne ed interna) e delle relative varianze (rapportando le devianze e ai relativi gradi di libertà).

I risultati dell'ANOVA a due criteri di classificazione per i master di primo livello possono essere riassunti come segue:

Fonte di variazione	Devianza (SQ)	Gradi di libertà	Varianze (MQ)	F empirici
fattore a (master)	48,671	2	24,336	F(a) = 0.820
fattore b (pubblicità)	194,657	2	97,329	F(b) = 3.281
fattore e	118,672	4	29,668	
TOTALE	362	8		

Dato che i gradi di libertà sia per il fattore *a* che per il fattore *b* sono pari a 2, i valori empirici di F, risultanti dai calcoli statistici, dovranno essere confrontati con lo stesso valore atteso sulla base della distribuzione F individuata dai corrispondenti gradi di libertà (2, 4) e con un livello di significatività pari a 0,01 e a 0,05:

$$F_{(0,05; 2; 4)} = 6,94$$

$$F_{(0,01; 2; 4)} = 18,00$$

Dal confronto risulta che:

$$F(a) < F_{(0,05; 2; 4)} \quad \text{accetto l'ipotesi } H_0 \quad F(a) < F_{(0,01; 2; 4)} \quad \text{accetto l'ipotesi } H_0$$

$$F(b) < F_{(0,05; 2; 4)} \quad \text{accetto l'ipotesi } H_0 \quad F(b) < F_{(0,01; 2; 4)} \quad \text{accetto l'ipotesi } H_0$$

I risultati del test F permettono di accettare, per entrambi i livelli di significatività, entrambe le ipotesi nulle: sarebbe possibile concludere che non esistono differenze significative tra le medie dei vari canali informativi e tra quelle dei master di I livello considerati. Molto probabilmente, però, questa conclusione non è dovuta all'assenza totale di differenze, ma al fatto che facendo una distinzione tra master di primo e di secondo livello il campione considerato non è più rappresentativo dell'universo considerato.

Master di II livello

Facendo riferimento ai dati relativi ai soli master di secondo livello (tab. 3.11) e procedendo con analisi statistiche analoghe a quelle viste in precedenza, i risultati dell'ANOVA possono essere così schematizzati:

Fonte di variazione	Devianza (SQ)	Gradi di libertà	Varianze (MQ)	F empirici
fattore a (master)	1878,4	5	375,680	F(a) = 3.268
fattore b (pubblicità)	955,05	2	477,525	F(b) = 4,154
fattore e	1149,661	10	114,966	
TOTALE	3983,111	17		

Per saggiare le ipotesi H_0 si considerano i seguenti valori della distribuzione F con un livello di significatività pari a 0,05:

$$F(a) = 3,268 \quad \text{confrontato con} \quad F_{(0,05; 5; 10)} = 3,33$$

$$F(b) = 4,154 \quad \text{confrontato con} \quad F_{(0,05; 2; 10)} = 4,10$$

per cui

$$F(a) < F_{(0,05; 5; 10)} \quad \text{accetto l'ipotesi } H_0$$

$$F(b) > F_{(0,05; 2; 10)} \quad \text{rifiuto l'ipotesi } H_0$$

Passando da un livello di significatività pari a 0,05 ad uno pari a 0,01 i valori empirici devono essere confrontati con i seguenti valori teorici:

$$F_{(0,01; 5; 10)} = 5,64 \quad \text{quindi} \quad F(a) < F_{(0,01; 5; 10)} \quad \text{accetto l'ipotesi } H_0$$

$$F_{(0,01; 2; 10)} = 7,56 \quad \text{quindi} \quad F(b) < F_{(0,01; 2; 10)} \quad \text{accetto l'ipotesi } H_0$$

Per i master di II livello i risultati dell'ANOVA ci portano ad affermare che, qualunque sia il margine d'errore concesso, non esistono differenze significative tra i diversi master, quindi la tipologia di master non influenza il

comportamento di coloro che presentano domanda di iscrizione. Per quanto riguarda, invece, il canale informativo utilizzato è possibile riscontrare differenze tra le medie solo se si è disposti ad accettare un margine di errore pari al 5%. Se, invece, consideriamo un livello di significatività pari a 0,01 siamo costretti ad accettare l'ipotesi H_0 (di uguaglianza tra le medie) e quindi anche il canale informativo non sarà influente sulla scelta del master al momento dell'iscrizione. Anche in questo caso, comunque, tali conclusioni potrebbero essere dovute al fatto che facendo una distinzione tra master di primo e di secondo livello il campione considerato perde di rappresentatività dell'universo.

Tuttavia, per i master di secondo livello e con la probabilità di commettere un errore pari al 5%, alcune differenze possono essere riscontrate tra le medie dei canali informativi considerati. Per individuare quali medie si discostano dalle altre è necessario utilizzare il metodo dei contrasti ortogonali.

Un primo confronto, tra la media del canale “internet” e gli altri due canali informativi, può essere così formulato:

$$l_1 = 2 * 24,333 - 13,333 - 6,667 = 28,666$$

da cui l’F empirico sarà:

$$F_1 = (28,6662 / (4/6 + 1/6 + 1/6)) / 114,9961 = 7,148$$

Il secondo confronto deve essere fatto fra le medie dei canali informativi “stampa” e “altro”:

$$l_2 = 13,333 - 6,667 = 6,666$$

e il relativo F sarà:

$$F_2 = (6,6662 / (1/6 + 1/6)) / 114,9961 = 1,59$$

Il valore critico della distribuzione F con cui F_1 ed F_2 dovranno essere confrontati è il seguente:

$$F_{(0,05; 1; 10)} = 4,96$$

Dal confronto risulta quanto segue:

$F_1 > F_{(0,05; 1; 10)}$ rifiuto l'ipotesi H_0 di uguaglianza tra le medie

$F_2 < F_{(0,05; 1; 10)}$ accetto l'ipotesi H_0 di uguaglianza tra le medie

Con un livello di significatività pari a 0,05, la media del canale informativo internet si scosta dalle altre, mentre le medie di stampa e altro non mostrano differenze significative. Anche in questo caso internet risulta essere il mezzo di comunicazione più efficace.

Tuttavia, per ridurre i margini di errore ed avere dei risultati più significativi sarebbe opportuno aumentare la numerosità dei campioni considerati, in modo da renderli il più possibile rappresentativi dell'universo di coloro che presentano domanda di iscrizione ai master distinti tra master di primo e di secondo livello.

Ai fini dell'analisi qui condotta, potrebbe essere utile valutare per ciascun master il canale informativo più efficace nella comunicazione con i candidati all'iscrizione ad un MasterDEA. Tuttavia, come risulta dalle analisi precedenti, la semplice distinzione tra master di primo e di secondo livello, ha portato ad una perdita di rappresentatività del campione considerato e a maggior ragione

le analisi condotte sullo stesso campione suddiviso per ciascun master potrebbero portare a conclusioni analoghe. Di conseguenza, per giungere a risultati significativi, sarebbe opportuno aumentare la numerosità del campione ed analizzarlo con tecniche statistiche simili a quelle precedenti o, comunque, adeguate al tipo di informazioni di cui si dispone.

3.6.5

L'analisi di correlazione

Dall'ANOVA condotta su tutto il campione, senza fare alcun tipo di distinzione, (vedi paragrafo 3.6.3) è risultato che “internet”, rispetto agli altri due canali informativi, è il mezzo più efficace nella comunicazione con i soggetti che sono alla ricerca di informazioni sull'offerta formativa post-laurea.

Ciò premesso, potrebbe essere utile analizzare l'esistenza o meno di legami statistici tra il numero di accessi ai siti MasterDEA che sono stati effettuati nel periodo di campagna pubblicitaria (x_i) (vedi tab. 3.6) e le frequenze rilevate dai questionari di coloro che presentano domanda di iscrizione dichiarando di esser venuti a conoscenza dei suddetti master da internet (y_i). L'obiettivo è quello di cogliere a pieno non solo l'efficacia, ma anche l'efficienza di questo canale che si va sempre più affermando come mezzo di comunicazione privilegiato nella logica di marketing.

La discussione che si sta sviluppando, mira alla costruzione di uno specifico indice statistico per lo studio delle relazioni tra le componenti della variabile doppia $(x_i; y_i)$ associata a ciascun master. L'analisi comincia dalla rappresentazione grafica dei risultati ottenuti dalla rilevazione delle due variabili per le stesse unità statistiche (tab. 3.12), allo scopo di avere un'idea circa l'eventuale legame che potrebbe sussistere tra tali variabili.

MASTER	Visite on line (x_i)	Domande internet (y_i)
BAA	6872	10
EAM	9004	19
FCG	8575	28
MANSAN	4712	15
GSA	7532	9
ACI	21691	60
AFC	4045	25
FCG_OL	20110	25
TOTALE	82541	191
Valori medi	10317.5	23.875

Tabella 3.12. Frequenze di visita on line (x_i) e di domande le cui informazioni sono state ottenute da internet (y_i).

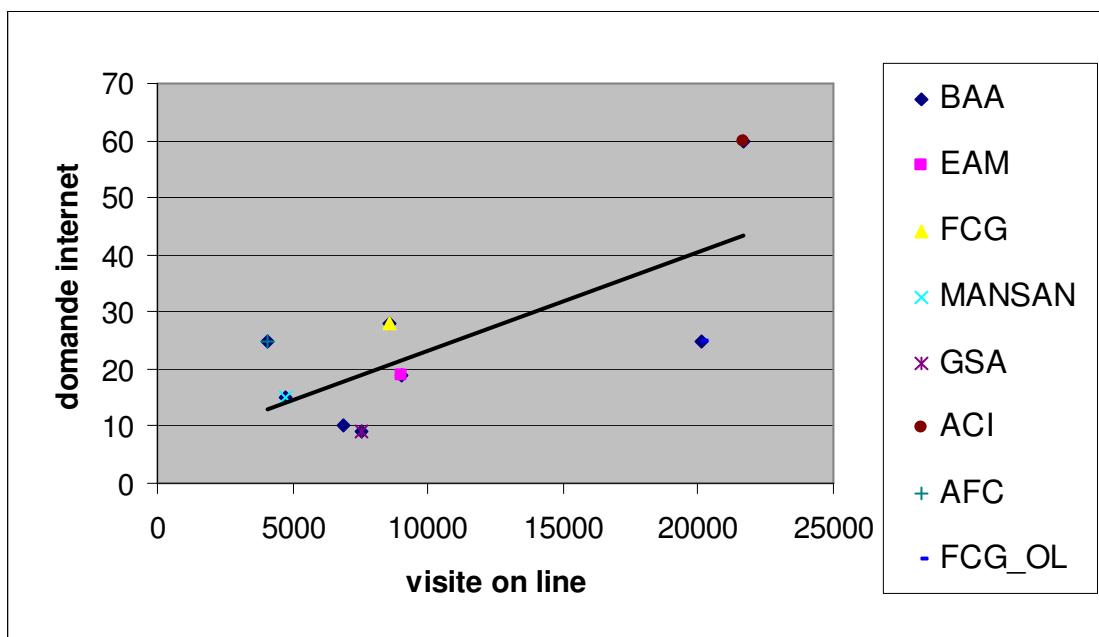


Figura 3.13. Illustrazione grafica dei dati in tabella 3.12.

I dati riportati in grafico (fig. 3.13) evidenziano variazioni congiunte delle due variabili. Il problema è quindi quello di esprimere tale associazione con un indice quantitativo.

Il metodo statistico adatto alla misura del grado e del tipo di interdipendenza tra x ed y è il **coefficiente di correlazione semplice lineare** di Bravais-Pearson:

$$r_{xy} = \frac{S_{xy}}{S_x * S_y}$$

dove

- S_{xy} è la **covarianza**;
- S_x e S_y sono le **deviazioni standard** di x e di y .

La tabella che segue esplicita i calcoli necessari per la determinazione dell'indice di correlazione.

MASTER	$X_i - X_{\text{medio}}$	$y_i - y_{\text{medio}}$	$(X_i - X_{\text{medio}})^2$	$(y_i - y_{\text{medio}})^2$
BAA	-3445,625	-13,875	11872331,64	192,515625
EAM	-1313,625	-4,875	1725610,641	23,765625
FCG	-1742,625	4,125	3036741,891	17,015625
MANSAN	-5605,625	-8,875	31423031,64	78,765625
GSA	-2785,625	-14,875	7759706,641	221,265625
ACI	11373,375	36,125	129353658,9	1305,015625
AFC	-6272,625	1,125	39345824,39	1,265625
FCG_OL	9792,375	1,125	95890608,14	1,265625
TOTALE			320407513,9	1840,875

Si procede con il calcolo del numeratore (covarianza):

$$S_{xy} = \frac{\sum_i (x_i - x_{medio}) (y_i - y_{medio})}{N} = 6,685$$

Al denominatore, invece, c'è il prodotto tra le deviazioni standard di x e di y.

Dunque sarà necessario calcolare preliminarmente le varianze:

$$S^2_x = \frac{\sum_i (x_i - x_{medio})^2}{N} = 3872,837$$

$$S^2_y = \frac{\sum_i (y_i - y_{medio})^2}{N} = 0,022$$

e poi la deviazioni standard:

$$S_x = \sqrt{S^2_x} = 62,232$$

$$S_y = \sqrt{S^2_y} = 0,149$$

Solo ora si hanno a disposizione tutti i dati per calcolare il coefficiente di correlazione:

$$r_{xy} = \frac{S_{xy}}{S_x * S_y} = \frac{6,685}{62,232 * 0,149} = 0,72$$

I dati evidenziano una consistente relazione positiva tra le visite on line e le domande presentate, cioè: al crescere del numero delle visite cresce anche il numero di soggetti che presentano domanda di ammissione ad uno dei MasterDEA. Questa analisi conferma ancora una volta la validità del canale

informativo “internet” quale mezzo pubblicitario principe nella diffusione delle informazioni relative all’offerta formativa post-laurea del Dipartimento di Economia Aziendale dell’Università di Pisa. È proprio su questo canale che potrebbero essere indirizzati i maggiori sforzi economici al fine di migliorare e massimizzare il profitto derivante dall’azione pubblicitaria.

3.6.6

L’analisi finanziaria

Alla luce delle analisi statistiche sopra riportate, la misurazione dei risultati finanziari è un passo necessario per poter compiere una valutazione dell’efficienza (in termini di costi sostenuti e di risultati conseguiti) della campagna pubblicitaria on-line tra i vari master.

Il linea di principio, l’analisi costi-benefici viene usata nella valutazione di progetti di investimento e costituisce un valido strumento per guidare le scelte tra ipotesi alternative. Il progetto rappresenta la cornice al cui interno si manifesteranno i benefici e i costi dei quali si tenterà la valutazione.

Nel caso in esame, il risultato dei progetti di investimento pubblicitario può essere adeguatamente misurato attraverso il valore dei profitti ottenuti dall’investimento stesso. Opportuni confronti tra i “benefici” e i “costi” sostenuti, permetteranno di decidere quale tra le campagne pubblicitarie on-line è risultata più proficua. Ovviamente, costi e benefici possono essere definiti soltanto in relazione agli obiettivi prestabiliti e, quindi, alle variazioni

di visite che l'azione pubblicitaria ha prodotto e al numero di soggetti che hanno presentato domanda di iscrizione dopo aver visitato i siti MasterDEA.

Al fine della valutazione costi-benefici relativa agli investimenti pubblicitari on-line non sono stati presi in considerazione i due master che sono alla loro prima edizione (AFC e FCG_OL) perché su questi è impossibile quantificare l'effetto in termini di variazione percentuale tra il prima e il dopo rispetto alla campagna pubblicitaria. Per i restanti master, qui di seguito riportati (tab. 3.14), è stata preliminarmente calcolata l'incidenza che il canale "internet" ha avuto sul totale delle domande costituenti il campione considerato.

MASTER	Internet	TOTALE	Incidenza internet
BAA	10	26	38,46%
EAM	19	32	59,38%
FCG	28	37	75,68%
MANSAN	15	27	55,56%
GSA	9	20	45,00%
ACI	60	100	60,00%

Tabella 3.14. Incidenza di internet sul totale delle domande campione.

Per benefici e costi di un progetto di investimento si intendono rispettivamente i guadagni (monetari e non) ottenuti dalla realizzazione del progetto stesso e le risorse consumate nel corso di tale realizzazione. Nel caso in esame si è proceduto alla quantificazione monetaria degli effetti della pubblicità on-line prendendo in considerazione i ricavi totali ottenibili solo se il numero dei discenti iscritti fosse pari al numero degli allievi ammessi (tab. 3.15).

MASTER	Allievi ammessi	Prezzo Master		Ricavi Totali
BAA	35	€ 4.800,00		€ 168.000,00
EAM:				€ 154.500,00
Full Time	15	€ 5.500,00	€ 82.500,00	
Part Time	15	€ 4.800,00	€ 72.000,00	
FCG	30	€ 5.800,00		€ 174.000,00
MANSAN	25	€ 5.800,00		€ 145.000,00
GSA	30	€ 5.800,00		€ 174.000,00
ACI:				€ 687.000,00
Auditing & Governance	10	€ 6.800,00	€ 68.000,00	
Finanza e controllo aziendale	10	€ 6.800,00	€ 68.000,00	
Auditing & Consulting	10	€ 5.800,00	€ 58.000,00	
Controllo di Gestione	10	€ 5.800,00	€ 58.000,00	
ACI Part Time	75	€ 5.800,00	€ 435.000,00	

Tabella 3.15. Ricavi totali MasterDEA.

Per monetizzare i benefici che internet ha apportato ai ricavi totali è necessario calcolare la parte di ricavi relativa ai soggetti che hanno presentato domanda di iscrizione dichiarando di aver conosciuto l'offerta formativa MasterDEA dal web (tab. 3.16). Dal prodotto tra i ricavi totali e l'incidenza di internet risulta quanto segue:

MASTER	Ricavi Totali	Incidenza Internet	Benefici internet
BAA	€ 168.000,00	38,46%	€ 64.612,80
EAM	€ 154.500,00	59,38%	€ 91.742,10
FCG	€ 174.000,00	75,68%	€ 131.683,20
MANSAN	€ 145.000,00	55,56%	€ 80.562,00
GSA	€ 174.000,00	45,00%	€ 78.300,00
ACI	€ 687.000,00	60,00%	€ 412.200,00

Tabella 3.16. Calcolo benefici internet.

Successivamente, per il calcolo dei benefici ottenuti per l'a.a. 2006/2007 dall'azione pubblicitaria on-line (tab. 3.17), sono state prese in considerazione le variazioni percentuali delle visite al sito che sono intervenute tra il prima e il dopo la campagna di comunicazione (vedi paragrafo 3.6.1, tab. 3.6-3.7).

MASTER	Benefici internet	Variazione % visite on line	Benefici campagna pubblicitaria on-line
BAA	€ 64.612,80	149,08%	€ 96.324,76
EAM	€ 91.742,10	104,40%	€ 95.778,75
FCG	€ 131.683,20	104,02%	€ 136.976,86
MANSAN	€ 80.562,00	36,82%	€ 29.662,93
GSA	€ 78.300,00	62,54%	€ 48.968,82
ACI	€ 412.200,00	52,79%	€ 217.600,38

Tabella 3.17. Calcolo benefici campagna pubblicitaria on-line.

Solo adesso sarà possibile mettere a confronto i benefici ottenuti con i relativi costi. I progetti di investimento pubblicitario possono essere confrontati attraverso la differenza tra i benefici e i costi ($B - C$) e l'analisi del cosiddetto rapporto benefici-costi (B / C). Mentre la prima è utile a mettere in evidenza il valore assoluto di un progetto, con il rapporto B / C si è in grado di evidenziare i benefici ottenuti per unità di capitale investito: se il rapporto è maggiore di uno, il progetto può essere accolto⁴. Tra i vari investimenti pubblicitari on-line, ovviamente, potrà essere considerato più efficiente quello che presenta il rapporto più elevato.

Nella tabella che segue (tab. 3.18) si mettono a confronto i benefici sopra stimati e i costi risultanti dalle fatture di acquisto dei servizi on-line dai vari operatori di settore (fonte MasterDEA).

MASTER	Benefici campagna pubblicitaria on-line (B)	Costi sostenuti (C)	B - C	B / C
BAA	€ 96.324,76	€ 5.500,00	€ 90.824,76	17,514
EAM	€ 95.778,75	€ 8.740,00	€ 87.038,75	10,959
FCG	€ 136.976,86	€ 8.080,00	€ 128.896,86	16,953
MANSAN	€ 29.662,93	€ 1.200,00	€ 28.462,93	24,719
GSA	€ 48.968,82	€ 6.890,00	€ 42.078,82	7,107
ACI	€ 217.600,38	€ 9.000,00	€ 208.600,38	24,178

Tabella 3.18. Analisi costi – benefici.

⁴ Nuti F., *L'analisi costi-benefici*, Il Mulino, Bologna, 1987.

Dall'analisi dei risultati, emerge chiaramente che tutti i progetti di investimento pubblicitario on-line sono convenienti (tutte le differenze benefici – costi sono maggiori di zero), ma la maggiore efficienza in termini di ritorno sull'investimento riguarda i master MANSAN e ACI. Ciò significa che, per questi master, i benefici per unità di capitale investito sono molto più elevati rispetto ad altri come GSA ed EAM.

Da analisi di questo tipo possono essere tratte alcune indicazioni utili al miglioramento dell'allocazione del budget pubblicitario. Ad esempio si potrebbe ipotizzare una riduzione dell'investimento on-line per GSA perché, forse, i soggetti interessati a questo master prediligono altri mezzi di comunicazione per ottenere informazioni. Questa supposizione, comunque, dovrebbe trovare conferma in analisi più approfondite sulla differenziazione dei canali informativi per ciascun master, che in questa sede non è stato possibile condurre per mancanza di rilevazioni a riguardo.

4

Conclusioni

4.1

I risultati dell'analisi

Per l'analisi del caso "MasterDEA" è stato preso in considerazione un campione di domande presentate, nell'a.a. 2006/2007, per l'ammissione ai master, considerando la sola parte relativa ai canali informativi attraverso cui gli allievi sono venuti a conoscenza dell'offerta formativa post-laurea del Dipartimento di Economia Aziendale dell'Università di Pisa.

Sono state anche analizzate le frequenze di visita dei siti internet pubblicati per ciascun master per vedere se e in che modo la pubblicità influenza i comportamenti di coloro che sono alla ricerca di informazioni su ciò che il mercato della formazione offre.

Tali dati, raccolti ed aggregati secondo logiche di significatività statistica, sono stati analizzati al fine di ottenere risultati che potessero in qualche modo sostenere le strategie di investimento pubblicitario fin ora utilizzate dall'area marketing MasterDEA.

Anzitutto sono state analizzate le variazioni che durante il periodo considerato (dal 01/02/2006 al 31/01/2007) sono intervenute nel numero di visite medie giornaliere. È stato riscontrato un notevole aumento del numero degli accessi on-line soprattutto nei periodi in cui entrano in atto le campagne pubblicitarie, periodi che quasi sempre sono antecedenti alle date di scadenza dei bandi di ammissione. Da ciò è stato possibile dedurre che la pubblicità è fonte preziosa di comunicazione anche in un settore, quale quello della formazione, che negli ultimi anni ha visto lo sviluppo notevole della concorrenza.

Sulla base di queste considerazioni iniziali si è proceduto con un'analisi "prima-dopo" per la valutazione degli effetti della pubblicità sui comportamenti di visita on-line. I risultati mostrano che prima della campagna pubblicitaria le visite on-line tra i vari master sono più omogenee, nel senso che i siti vengono visitati in modo indifferente da soggetti in cerca di informazioni generali. La pubblicità ha, invece, l'effetto di indirizzare i visitatori verso i siti internet di loro interesse facendone aumentare le visite e rendendo più eterogenei i comportamenti di visita soprattutto nei periodi antecedenti la data di scadenza in cui ricadono le maggiori attività pubblicitarie.

Data la validità dell'azione pubblicitaria, l'attenzione è stata rivolta all'individuazione di connessioni tra i canali informativi al fine di stabilire se ha senso differenziare la pubblicità o è meglio focalizzarsi su uno dei canali informativi. A tale scopo le risultanze delle domande campione sono state raggruppate in tre macro-classi ("internet", "stampa" ed "altro") per avere maggiore uniformità di risposta. Dal calcolo dell'indice quadratico di connessione risulta che le tre categorie di pubblicità sono strettamente connesse e che l'una non può prescindere dalle altre. Tutte, quindi, concorrono a far conoscere ai soggetti interessati l'offerta formativa post-laurea MasterDEA.

Definita la validità di tutti i mezzi di comunicazione utilizzati, qual è, però, quello più efficace? Per rispondere è stato necessario scendere più in profondità con le analisi per vedere se esistono delle differenze significative tra le macro-classi considerate. Con una probabilità molto bassa di commettere un errore, l'ANOVA a due criteri di classificazione (canali informativo – master) e i relativi test di verifica delle ipotesi si sono rivelati strumenti validi per mettere in evidenza che esistono differenze significative tra i canali informativi, ma che la tipologia di master non ha influenze sul comportamento di colui che presenta domanda di iscrizione. Ciò significa che un soggetto presenta domanda di iscrizione ad uno dei MasterDEA non perché cerca quel tipo di master, ma perché la pubblicità gli ha fatto conoscere ciò che offre il Dipartimento di Economia Aziendale dell'Università di Pisa.

Ma per conoscere quale tra le pubblicità ha un maggior *appeal* su colui che è alla ricerca di informazioni è stato necessario mettere a confronto le medie delle tre macro-classi pubblicitarie individuate e vedere quale effettivamente si scosta dalle altre. Il metodo dei contrasti ortogonali ha messo in luce una differenza significativa tra la media del canale internet e quelle delle altre due macro-classi individuate. È stato possibile concludere che “internet” rispetto alla “stampa” e ad “altri” è il mezzo di comunicazione più efficace. D'altronde le possibilità oggi offerte dall'interattività e dalla multimedialità della rete web fanno di internet uno straordinario mezzo di comunicazione pubblicitaria e non, consentendo a chiunque di avere in tempo reale un'ampia gamma di informazioni che difficilmente potrebbero essere fornite con un manifesto o con una pubblicità su carta stampata.

Con analisi statistiche simili a quelle precedenti si è cercato di mettere in luce se sia possibile utilizzare diversi mezzi pubblicitari per master differenti. Una prima distinzione logica e coerente con la struttura organizzativa dei

MasterDEA è quella tra master di primo e di secondo livello. Per i primi è risultato che non ci sono differenze significative né tra i canali informativi, né tra i master facenti parte del sottogruppo individuato. Tali risultati sono contrastanti con quanto visto in precedenza (master considerati senza alcuna distinzione di sorta) e quindi molto probabilmente non sono dovuti all'assenza totale di differenze, ma al fatto che facendo una distinzione tra master di primo e di secondo livello il campione non è più rappresentativo dell'universo considerato. Con riferimento ai soli master di secondo livello, invece, le analisi evidenziano che non esistono differenze significative tra i diversi master; mentre è possibile affermare che i canali informativi influiscono sui comportamenti di iscrizione solo se si è disposti ad accettare un aumento della probabilità di errore. Questa è un'ulteriore conferma della perdita di rappresentatività del campione considerato dopo la distinzione tra master di primo e di secondo livello. Anche in questo caso con il metodo dei contrasti ortogonali, si è constatato statisticamente la rilevanza di “internet” quale mezzo di comunicazione più efficace nel pubblicizzare l'offerta formativa post-laurea MasterDEA.

Stabilita la validità pubblicitaria del canale informativo “internet”, altro obiettivo del presente lavoro è stato quello di valutarne l'efficienza. A tale scopo è stato utile indagare sull'esistenza di correlazioni tra gli accessi ai siti MasterDEA e il numero di domande di iscrizione in cui è stato dichiarato di aver utilizzato internet per conoscere l'offerta formativa del Dipartimento di Economia Aziendale: l'analisi evidenzia una forte correlazione positiva. Ciò sta ad indicare che ad un aumento del numero di visite on-line corrisponde un incremento del numero di domande presentate. Internet, ancora una volta, può essere considerato il canale informativo da privilegiare nella pianificazione della strategia pubblicitaria.

Sempre nel tentativo di esprimere un giudizio sulla campagna di comunicazione MasterDEA per l'a.a. 2006/2007, si è pensato di approfondire l'indagine anche sul versante finanziario. Con l'analisi costi-benefici è stata valutata la convenienza del progetto di investimento pubblicitario on-line. Da tale analisi è risultato che per tutti i master i benefici ottenuti sono maggiori dei costi, quindi le campagne pubblicitarie hanno degli effetti positivi in termini di ritorni sui costi sostenuti. Mentre analizzando il rapporto benefici-costi (B / C), è possibile notare come le campagne pubblicitarie on-line per ACI e MANSAN sono molto più efficienti. Ciò significa che, per questi master, i benefici per unità di capitale investito è molto più elevata rispetto ad altri master come GSA ed EAM.

Per giungere ad una migliore allocazione del budget pubblicitario, comunque, sarebbe stato utile estendere l'analisi a tutti i canali informativi. In questa sede, però, non è stato possibile condurre tale verifica per mancanza di rilevazioni a riguardo.

In estrema sintesi ed in modo oggettivo, dai risultati delle analisi statistiche condotte è possibile trarre le seguenti considerazioni:

- la pubblicità on-line è di fondamentale importanza per pubblicizzare l'offerta formativa post-laurea MasterDEA;
- gli investimenti pubblicitari sostenuti per la pubblicità on-line sono totalmente giustificati dai risultati in termini di contatti ottenuti;
- gli altri canali informativi, stampa ed altro, fanno da cornice ad internet e possono essere considerati come input necessari affinché un soggetto venga a conoscenza dei master e vada in cerca di maggiori informazioni utilizzando le risorse on-line.

4.2

Considerazioni

La statistica è una metodologia, cioè un insieme di procedure e di strumenti per l'elaborazione di dati utili per la conoscenza e la soluzione di problemi scientifici e pratici. Per queste sue caratteristiche la statistica offre uno “strumentario” che esperti di altre discipline possono utilmente impiegare per risolvere i loro problemi.

In questo lavoro di ricerca si è cercato di delineare un panorama delle più elementari e diffuse applicazioni statistiche per una miglior conoscenza dei fenomeni aziendali ed approfondire l'utilità di questi strumenti nelle attività di marketing che sono parte integrante ed indispensabile dell'area commerciale di un'azienda.

L'analisi statistica è oggi il valore aggiunto di un gran numero di imprese di successo, in quanto consente loro di analizzare i dati e di trarre conclusioni. Invece di affidarsi alle sensazioni, i dirigenti quantificano le relazioni utilizzando l'analisi statistica e, in conseguenza di ciò, sono in grado di prendere decisioni migliori. Pianificazione, analisi, valutazione, gestione delle risorse, controllo delle performance diventano molto più efficaci se gestite con strumenti statistici.

L'analisi statistica permette a chi la utilizza di trovare soluzioni a problemi in una varietà di settori. La panoramica delle principali applicazioni statistiche, fornita in questo lavoro, è finalizzata ad una migliore conoscenza dei fenomeni aziendali attraverso tecniche di *Data Mining*, e mette in evidenza il valore aggiunto che una corretta analisi critica dei dati statistici può fornire.

Nella tesi sono stati analizzati vari campi di applicazione di tali sistemi di business intelligence rivolti al CRM (*Customer Relationship Management*), cioè alla gestione efficace dei rapporti con la clientela; è importante puntualizzare però che i sistemi di business intelligence possono e dovrebbero essere diffusi in tutte le aree dell'organizzazione. Tale precisazione è essenziale perché l'applicazione di questi sistemi avviene solitamente solo in alcune aree funzionali (ad es. produzione e vendite) e perché spesso essi vengono applicati solo su un sottoinsieme dei problemi decisionali di competenza della stessa area; in generale però non vi è a priori nessuna ragione o vincolo affinché le tecnologie di business intelligence non possano essere applicate all'azienda nella sua totalità.

Si è cercato, inoltre, di evidenziare in maniera critica ed analitica quale importante contributo possa fornire la statistica alla gestione di un' "azienda" operante nel settore della formazione post-laurea, come il Dipartimento di Economia Aziendale dell'Università di Pisa, per la gestione dei contatti con i soggetti interessati a tale offerta formativa. L'obiettivo principale è stato quello di valutare l'efficacia e l'efficienza dell'azione di comunicazione pubblicitaria MasterDEA.

L'analisi ha, comunque, presentato alcune difficoltà soprattutto in termini di disponibilità delle informazioni e per due ordini di motivi. In primo luogo, nell'elaborazione dei dati raccolti si è cercato con analisi più approfondite, di individuare possibilità di diversificazione dei canali informativi in funzione del master considerato. Ciò non è stato possibile perché la numerosità dei campioni relativi ai singoli master perdevano di rappresentatività se considerate separatamente.

In secondo luogo non è stato possibile indagare sulla validità delle campagne pubblicitarie effettuate su stampa o su altri mezzi di comunicazione perché non si disponeva delle informazioni relative alla tiratura dei giornali pubblicati, al quantitativo di affissioni, al numero di passaggi del messaggio pubblicitario

per radio, ecc. Queste informazioni sarebbero, infatti, state utili per una comparazione tra gli sforzi pubblicitari e gli effetti da questi ottenuti così come è stato possibile fare con “internet” di cui si disponeva del numero degli accessi.

Lungo tutto il corso della trattazione si è voluto mettere in evidenza (soprattutto alla luce delle analisi condotte sul caso MasterDEA) che non è corretto affermare, come spesso si sente nelle aziende, che la statistica è inutile nel processo decisionale, né che mancano strumenti statistici da adoperare. Gli strumenti ci sono, basta capire in quale contesto e come utilizzarli. Le analisi qui condotte ne sono una dimostrazione evidente.

Va comunque sottolineato che quelli presentati in questo lavoro sono solo alcuni degli strumenti della materia statistica applicata alla gestione manageriale e che con competenze evolute a riguardo si potrebbero sviluppare analisi e controlli maggiormente precisi e produttivi.

Per un manager/imprenditore aiutare il proprio intuito con la statistica può voler significare dare alla propria azienda il massimo dei risultati in termini di gestione delle politiche produttive, fluidità del circolante interno, precisione nel raggiungimento dei *budget*, attenzione nel controllo delle vendite, politiche di *marketing* e di relazione con il cliente e non ultimo incremento delle *performance* e della motivazione del proprio *team*.

BIBLIOGRAFIA

ALBANO A., *Costruire sistemi per basi di dati*, Addison Wesley Longman Italia, Milano, 2001.

BERRY M., LINOFF G., *Data mining techniques: for marketing, sales and customer relationship management*, Wiley, New York, 1997.

BERSON A., SMITH S. J., *Data Warehousing, Data Mining and OLAP*, McGraw-Hill, Milano, 1997.

BRASINI S. et al., *Marketing e pubblicità*, Il Mulino, Bologna, 1999.

BRASINI S. et al., *Statistica aziendale e analisi di mercato*, Il Mulino, Bologna, 2002.

CALORI G., *Dai sistemi di ERP alla business intelligence*, in “Amministrazione & Finanza”, 1, 2003, pp. 3-16, IPSOA, Milano.

CICCHITELLI G., HERZEL A., MONTANARI G. E., *Il campionamento statistico*, Il Mulino, Bologna, 1997.

CODA R., *Un controllo di gestione di “Qualità totale”*, in “Amministrazione & Finanza”, 19, 1998, pp. 33-37, IPSOA, Milano.

COFFETTI, E., *Metodologie quantitative, oggi in azienda*, in “Economia & Management”, 1, 2001, pp. 40-41, SDA Bocconi, Milano.

D’AMBRA L., AMENTA P., *Un approccio statistico multidimensionale per la valutazione della performance dei servizi di interfaccia con informazioni esterne*, in “Qualità”, vol. C, 2000, pp. 13-28, XX Convegno nazionale dell’AICQ, Bologna.

DEL CIELLO N. et al., *Metodi di data mining per il customer relationship management*, Franco Angeli, Milano, 2000.

FIorentino A., *la teoria della soddisfazione del cliente e le sue implicazioni per le strategie di impresa*, in “Finanza, Marketing e Produzione”, 2, 1996, pp. 105-136, EGEA, Milano.

GINI C., *Statistica e induzione*, CLUEB, Bologna, 2001.

GIUDICI P., *Data mining: metodi statistici per le applicazioni aziendali*, McGraw-Hill, Milano, 2001.

GIUDICI P., CASTELO C., *Association models for web mining*, in “Journal of data mining and knowledge discovery”, 5, 2001, pp. 183-196, Springer Netherlands.

GREENBERG P., *CRM Customer Relationship Management*, APOGEO, Milano, 2001.

HAN J., KAMBER M., *Data Mining: Concepts and Techniques*, Morgan Kaufmann, San Francisco, 2000.

LAMBIN J.J., *Marketing*, Mc Graw-Hill, Milano, 1991.

LANDENNA G., *Fondamenti di statistica descrittiva*, Il Mulino, Bologna, 1994.

MANARESI A., *La relazione tra marche e clienti. Un'analisi multisettoriale*, Carocci, Roma, 1999.

MARIANI P., *La statistica in azienda*, Franco Angeli, Milano, 1991.

MOLTENI L., *Nuove frontiere per le ricerche di marketing: web research e web mining*, in "Economia & Management", 1, 2001, pp. 42-43, SDA Bocconi, Milano.

NUTI F., *L'analisi costi-benefici*, Il Mulino, Bologna, 1987.

Parr Rud O., *Data Mining Cookbook, modeling data for marketing, risk, and Customer Relationship Management*, Wiley, New York, 2000.

PASINI M., *I sistemi informativi direzionali. Le tecnologie dell'informazione a supporto dei processi manageriali d'azienda*, EGEA, Milano, 2004.

PICCOLO D., *Statistica*, Il Mulino, Bologna, 2000.

PICCOLO D., *Statistica per le decisioni*, Il Mulino, Bologna, 2004.

Poli M., *Data mining: tra mito e realtà*, in "Economia & Management", 1, 2001, pp. 36-37, SDA Bocconi, Milano.

RAIMONDI M. et al., *CRM analitico: il contributo della business intelligence*, in "Sistemi & Impresa", 9, 2003, pp. 75-77, ESTE, Milano.

RAJOLA F. (a cura di), *L'organizzazione dei sistemi di business intelligence nel settore finanziario*, Franco Angeli, Milano, 2000.

ROIGER R. J., *Introduzione al data mining*, Mc Graw-Hill, Milano, 2004.

TAMBURRANO E., *Il controllo statistico di qualità*, Franco Angeli, Milano, 1991.

TRINCA COLONEL R., *Data mining: come estrarre valore dai customer database*, in "Economia & Management", 6, 2002, pp. 46-49, SDA Bocconi, Milano.

VALDANI E., *L'impresa pro-attiva*, Mc Graw-Hill, Milano, 2000.

VANZANELLI D., *Data warehousing & customer relationship management: al di là del mito*, in "Sviluppo & Organizzazione", 184, 2001, pp. 94-97, SDA Bocconi, Milano.

ZANI S., *Analisi dei dati statistici, osservazioni in una e due dimensioni*, volume 1, Giuffrè, Milano, 1997.

ZANI S., *Analisi dei dati statistici, osservazioni multidimensionali*, volume 2, Giuffrè, Milano, 2000.

ZENGA M., *Metodi statistici per l'economia e l'impresa*, Giappichelli, Torino, 1994.